

In praise of Gates

The co-founder of Microsoft has made a mint from a business that many attack, but his efforts in Africa highlight a virtue: a philanthropic understanding of science. The world needs more of it.

"Nelson Mandela! Nel...son...Man...del...a...! Bill and Melinda! Bill...and...Mel...ind...a!" Not surprising that the man who led South Africa out of apartheid should be given a charged, hip-hop hero's welcome as he walked into a youth forum on HIV/AIDS in Hillbrow, the seediest district of Johannesburg. But Bill Gates, the Microsoft mogul everyone loves to hate, and his wife Melinda, the filthiest-rich couple on the planet? Here and in other poor parts of the world, they are quickly becoming Saint Bill and Miraculous Melinda, offering salvation from scourges such as malaria and AIDS, and the best thing to happen to Africa for a long time.

Improving global health is the aim of the Bill & Melinda Gates Foundation, which with a US\$25-billion endowment is the world's largest philanthropic fund. To see things at first hand, the couple also travelled to a research centre in rural Mozambique, announcing \$168 million in new malaria research grants, and to Botswana, to check on their \$50-million programme there to combat AIDS. They've pumped over \$3.2 billion into diseases neglected for too long.

The World Bank's World Development Report should be recommended reading for the rich and powerful. Bill Gates says he discovered the scale of the global health crisis on reading the 1993 report. "We thought other people were dealing with these problems, that the research dollars were being allocated. We are still in a state of shock that there is this vacuum here."

Gates now pores through the US Centers for Disease Control and Prevention's *Morbidity and Mortality Weekly Report*, and swots up on biology. He surrounds himself with scientists, and throws them research queries with an impish Woody Allenish inquisitiveness. His single-mindedness and hard-boiled business habits are a tonic in the bureaucratic world of international health.

Biggest bottlenecks

The foundation's funds are a pittance compared with the task, but they give Gates unparalleled leverage to shape global health, and to galvanize others to join the effort. The foundation's initial focus cannot be faulted, aiming to hit the biggest disease burdens, the most severe lack of funding, the biggest bottlenecks. The dilemma, as Gates acknowledges, is to choose between current and future needs.

In Botswana the AIDS programme is here and now. So is the Global Alliance for Vaccines and Immunization (GAVI) programme, set up in 1999 with \$750 million of Gates funds. In just three years it has delivered essential vaccines, such as for hepatitis B and yellow fever, to some 30 million children.

Such pilot projects show that distributing antiretrovirals in poor rural Africa is practical: over to governments to see the benefits and learn from best practice. Why Botswana? Because its government was sufficiently committed — other African governments, take note.

In creating GAVI, Gates didn't just buy vaccines but established a system for guaranteeing markets, pooling funds and thrashing out acceptable differential pricing deals for poorer countries with vaccine companies. It could be a model for others charged with getting existing or new tools out there.

Otherwise, Gates is right to leave the huge costs of funding control to governments — a proper attack on malaria alone would cost

\$2.5 billion a year. His main interest is the abyss between basic research and industrial development.

Take drug development for malaria, which had come to a standstill before the creation in 1999 of a global alliance, the Medicines for Malaria Venture. This programme is taking the best drug leads from academia, and pushing them through development to market. It has provided the biggest drug pipeline for malaria in three-quarters of a century. Shame that, as with most malaria-vaccine research and development, its funding largely depends on Gates.

The foundation, under Richard Klausner, ex-head of the US National Cancer Institute, should fund more upstream research. The sequencing of the malaria parasite genome showed much variation in the surface antigens used in most current candidate vaccines, for example. In the long run, better understanding of the parasite's biology, and its mosquito and human hosts, could prove most beneficial.

Blue-skies research

Gates has not funded much basic research. But the foundation recently announced a \$200-million Grand Challenge competition to identify blue-skies research ideas that could pay huge dividends in the fight against neglected diseases. Neglecting other strategic areas of basic research would be short-sighted. The foundation is discussing funding post-malaria genomics and the science of vector control. It could move both fields of endeavour forward faster.

Scientists can spur the foundation on if they club together in consortia to agree on big problems that need to be cracked, and roadmaps for how to go about it. They should then write to Gates (see www.gatesfoundation.org/Grants/EligibilityAndGuidelines).

It is encouraging that 80% of Gates funding is channelled through existing public-private alliances. Such alliances, bringing together funders, governments, UN agencies and private companies, have over the past decade emerged as the predominant organizational attack for a host of diseases, including polio, malaria, AIDS and tuberculosis.

But in philanthropy, personal experience counts. Gates's visit to the Manhica Health Research Centre in Mozambique opened his eyes to the importance of high-level research centres in logistically difficult areas where diseases are endemic.

Manhica, set up as recently as 1996, has gathered important baseline data on the 65,000 people living within 200 square kilometres, painstakingly recording mortality, morbidity, births, migration and other data. It is part of international networks trying to generate this basic information — the foundation for solid epidemiological studies to assess the impact of experimental interventions.

Gates seems surprised that it is possible to establish such high-level centres delivering health care and doing research. The goal is to learn from this and to get such centres replicated across Africa.

Now weighing in as a major force in global health, Bill Gates can allow himself a little less political correctness. In calling for the malaria problem to be made more visible, he let slip his gut feeling: "It's unfortunate in a way that, because of geography, malaria has been wiped out in the rich world." But he seems inhibited in discussing politics and public policy. Gates should use his weight to influence world leaders. It's time he put his mouth where his money is. ■





Route to success
Roadmap lays out future for US biomedical research
p438



Opening doors
Harvard's systems biology head sees vision become reality
p439



Getting together
Moves are afoot on a transnational research agency for Europe
p440



Demolition job
Indian minister quits over destruction of mosque
p442

Mosquito production mooted as fast track to malaria vaccine

Declan Butler, Paris

A leading US researcher is probing an audacious approach to develop a malaria vaccine by cultivating billions of parasites, irradiated to stop them causing disease, in swarms of live mosquitoes.

The approach, which builds on experimental results that are 35 years old, is viewed by other top malaria researchers with a mixture of curiosity, incredulity and admiration.

But its proponent, Stephen Hoffman, founder of Sanaria, a company based in Gaithersburg, Maryland, says he has become disillusioned with the prospects for the current main approach, which uses proteins found on the surface of *Plasmodium*, the parasite that carries the disease. The best of these vaccines has shown only short-term protection in 40–70% of recipients, indicating that a viable malaria vaccine is still years away.

So Hoffman is turning to 1967 results from Ruth Nussenzweig's group at New York University, who found that mice could be protected from malaria by being injected with a form of the parasite — known as a sporozoite — extracted from irradiated mosquitoes (R. S. Nussenzweig *et al.* **216**, 160–162; 1967). The sporozoites infect the liver in humans and provoke an immune response, but, because of the weakening effect of the radiation, they don't go on to infect red blood cells or cause malaria.

Individually exposing people to thousands of bites by live irradiated mosquitoes is hardly viable for large-scale immunization. But Hoffman claims it should be possible to produce billions of parasites to manufacture a vaccine directly. He intends to breed millions of mosquitoes, grow them on cultures of infected red blood cells, dissect out the sporozoites from the mosquitoes' salivary glands, purify and store them, and then find ways of getting them into people — all under the tough standards of the US Food and Drug Administration.

There is a basis for such a bold project — a re-analysis of published data by various research centres, including the Naval Medical Research Center and the Walter Reed Army



Infected mosquitoes bred in captivity could conceivably deliver an effective vaccine against malaria.



Stephen Hoffman:
back to basics.

Institute of Research, both in Silver Spring, Maryland. This shows that humans exposed to more than 1,000 bites by infected, irradiated mosquitoes develop over 90% protection against subsequent challenges, and that this protection lasts for more than 10 months (S. L. Hoffman *et al.* *J. Infect. Dis.* **185**, 1155–1164; 2002).

But the logistical problems are formidable. “The barriers

have seemed sufficiently daunting that no one has been willing to give it a try,” says Thomas Richie, director of clinical trials at the Naval Medical Research Center Malaria Program in Silver Spring.

“It’s a long shot,” says Adrian Hill, a malaria-vaccine expert at the University of Oxford, UK: “It’s worth a try, although the odds are heavily stacked against him.”

Hoffman says his work at Celera Genomics on the mosquito genome project helped to convince him that the obstacles could be overcome. He expects to raise \$3 million for the project, including a \$550,000 grant from the National Institutes of Health.

He will need to produce sufficient numbers of sporozoites, and to store them in a way that retains their ability to infect liver cells. A bite

from a mosquito injects around 10–20 sporozoites. So a minimum vaccine dose, equivalent to 1,000 bites, would require at least 10^4 – 10^5 sporozoites, to ensure that enough living parasites make it to the liver. This is already achievable from one mosquito, says Robert Sauerwein, head of medical parasitology at the University of Nijmegen in the Netherlands, a collaborator of Hoffman's. So one mosquito could produce one vaccine dose.

This assumes that most sporozoites will survive and be immunogenic after storage. If not, many more will be needed. And although mosquitoes inject sporozoites into the blood — an efficient route to the liver — intravenous vaccine injection is not an option. Hoffman thinks that vaccination under the skin should be feasible. But the route to the liver would have to be unusually efficient for this to work.

Hoffman says that with his initial funding he will make sporozoites of clinical-trial quality. If he is successful, a vaccine could be quickly tested in humans in areas of Africa where malaria transmission is high.

The plan is the “best show in town” for malaria vaccines, says Maurice Hilleman, director of the Merck Institute for Vaccinology in West Point, Pennsylvania. “It might even be the only show in town.” ■

A. D. HOFFMAN & R. C. THOMPSON

China plays cards close to its chest over manned space shot

David Cyranoski, Beijing

China is expected to attempt to send its first man into space later this month, sources say. But analysts have warned that the mission may be postponed by managers of China's military-run space programme.

A successful launch of the Shenzhou 5 space capsule would make China the third nation to send a human into orbit — and set up China's emerging space programme for a twenty-first-century rivalry with that of the United States.

Government officials and semi-official sources have suggested throughout the year that the flight is due to take place in the middle of this month, but no official date has been released. Phillip Clark, head of the UK-based Molniya Space Consultancy, speculates that China will announce the flight at very short notice and air it live on television.

But even basic details of the flight, such as the number of 'taikonauts' who will be on board, have remained closely guarded. An official at one Chinese space-research organization says that, for safety reasons, only a single person will ride in the three-seat launch vessel. "This will be a symbolic flight," the official says.

But Clark suggests that China might try to send more than one person into space, to get one-up on the initial, solo flights by the Soviet Union and the United States, four decades ago.

Many of China's space scientists and engineers consider the space programme to be a scientific one, even though it falls within the military budget. They want to see an end to US export-control rules that block collaboration with NASA and exclude China from the International Space Station project.

A success with Shenzhou 5 might move China towards its goal of having its own space station, observers of the programme say. The next step would be to carry out docking exercises between a Shenzhou craft and another orbital module.

China's manned space programme has been progressing slowly — it is already four years since the country launched its first unmanned space capsule. But time is not a pressing issue for the Chinese space programme, comments Brian Harvey, a space analyst based in Ireland. "It is a slow, determined and stable programme," he says. "They aren't racing against anybody." ■

NIH 'roadmap' charts course to tackle big research issues

Erika Check, Washington

The director of the US National Institutes of Health (NIH) has announced a long-term plan for the biomedical research agency in which its component institutes will join forces to tackle priority projects.

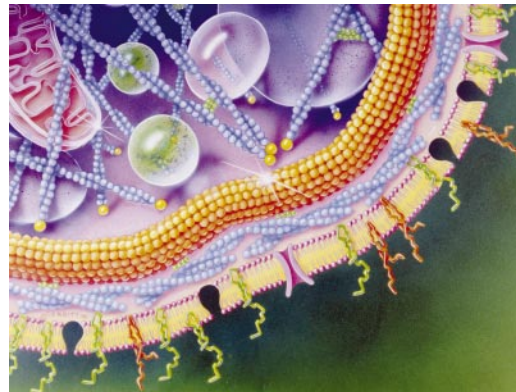
The NIH Roadmap, released by director Elias Zerhouni on 30 September, was developed by NIH institute directors and outside experts. Zerhouni says the NIH will begin implementing the plan in 2004, at an initial cost of \$125 million, rising rapidly to \$2 billion by 2009.

The plan includes 28 programmes in three main categories: new technologies, new types of research team and ways to improve clinical research infrastructure. NIH institutes will contribute from their individual budgets to support the programmes. "For the first time, there's been a real change in the way institutes fund common projects, because every institute has agreed to contribute to this common fund," Zerhouni says.

The plan will move the NIH along the road to supporting 'big' biology. For instance, the structural-biology programme calls for scientists to find new ways to produce large amounts of membrane-bound proteins and study them in a systematic way. Currently, there is no good way to isolate and study such proteins.

Another programme will set up libraries and databases of small molecules to be studied as potential drug treatments, and a third will create centres to write software for the analysis of biological data.

Another aim will be to build teams of investigators from diverse backgrounds: a series of programmes and conferences will



Membrane proteins are a key focus of the NIH's strategic plan.

foster collaborations between biologists and physical scientists. The roadmap also contains plans for the training of clinical researchers and for centres of excellence in translational research — the interface between lab work and patient care.

Zerhouni says the Roadmap is an attempt to address scientific problems that the NIH has struggled with in the past, such as multidisciplinary projects that require many institutes to work together. He acknowledged scientists' fears that biology research has shifted away from investigations by individuals, towards large, technology-driven science.

"Our sense from talking to the community is that we need to create new modes of exploration because the scale and complexity of scientific problems has increased," Zerhouni says. "That's what we're doing and we'll find out what the results are soon."

David Korn of the Association of American Medical Colleges says: "Universities, scientists and journals are all going to have to adjust to the multidisciplinary mode of science, and this plan moves in the right direction. But it's going to be a very unsettling change." ■

Ion-powered probe set for year-long Moon trek

Geoff Brumfiel

The first European spacecraft to visit the Moon was launched on 28 September. The craft is using an ion-powered thruster — an experimental propulsion system that the European Space Agency (ESA) hopes will carry probes to other planets.

The craft, called Small Missions for Advanced Research in Technology 1 (SMART-1), blasted off aboard an Ariane 5 rocket from French Guiana. It will now slowly spiral towards the Moon, without the use of typical chemical propellants.

When it arrives, SMART-1 will map the distributions of various elements on the lunar surface, says mission scientist Sarah Dunkin of the UK's Rutherford Appleton Laboratory: "It will help us distinguish between theories of the Moon's creation."

The ion thruster can only push the 367-kg craft with a force equivalent to a sheet of paper resting on an open hand. The journey will take a year and a quarter, rather than the few days taken by the Apollo missions. But given its bargain price of £100 million (US\$120 million), ESA is willing to wait. ■

Harvard heralds fresh take on systems biology

Erika Check, Washington

Harvard Medical School is to set up a department devoted to systems biology — the first entirely new department to be founded at the institution in 20 years.

Systems biology involves the integration of computer modelling, large-scale data analysis and biological experimentation. The 23 September announcement by the medical school at America's most famous university is seen as emblematic of the field's rapid rise — and the need to train a cadre of undergraduate and graduate students to work in it.

The department will be led by cell biologist Marc Kirschner, who pushed hard for its creation in the face of some resistance from established Harvard departments. Kirschner plans to recruit about 20 faculty members for the new department. Tim Mitchison of Harvard's cell biology department and Lewis Cantley of the Beth Israel Deaconess Medical Center have already decided to join the department. But Kirschner says that most of the rest will be recruited from outside the university.

Until now, most large US research universities, including the Massachusetts Institute of Technology and Stanford and Princeton universities, have sought to accommodate systems-biology expertise at interdisciplinary research institutes. Other centres, such as the Institute for Systems Biology in Seattle,



Teaching plan: Marc Kirschner campaigned for a facility to train students in quantitative biology.

Washington, have been set up independently, and yet other projects, such as the Alliance for Cellular Signaling, involve consortia of universities.

But by establishing a fully fledged department, Harvard hopes to incorporate undergraduate and graduate education into its work, says Joseph Martin, dean of the medical school.

The new department says it will cooperate closely with Harvard's Bauer Center for Genomics Research, set up just last year with

its own mission of bringing expertise from mathematics and the physical sciences to bear on biological problems.

Other systems-biology researchers say the challenge for the department will be to create career incentives for people who will be working without the boundaries of traditional disciplines. The systems-biology approach involves large teams of people, and it will be difficult to determine how best to make decisions about issues such as how to award tenure, suggests Leroy Hood, president of the Institute for Systems Biology.

Hood also points out that granting agencies, such as the National Science Foundation, are structured to support traditional research disciplines, and sometimes struggle to maintain support for large interdisciplinary groups. "There's got to be a mixture of people who make contributions," Hood says. "You may have an array facility or a proteomics facility, but you need an awful lot more than that to do integrated systems biology."

But Kirschner says this is precisely why he wanted to create a full department. "The alternative would be to create collaborative interactions among different groups," he explains. "But we want to create a community of people whose main interest is in bringing quantitative mathematical approaches to complex systems. I think it has the best chance of working if the group is created with this primary focus." ■

J. CLEARY/HARVARD MED. SCH.

Adviser gears up to bring scientists to matters of state

Geoff Brumfiel, Washington

George Atkinson's office window at the Department of State in Washington DC looks across the street to the National Academies, which advise the US government on scientific matters. But bringing the right technical expertise into the state department has been a perennial problem. As newly appointed science and technology adviser to the secretary of state, Atkinson is setting out to fix it.

He will take a small step towards that goal this week, when the department announces funding from the MacArthur Foundation, the Carnegie Foundation and a group of universities for a beefed-up scholarship programme that will pluck seasoned researchers from universities. After a year at the department, they will return to their jobs while also continuing as consultants with the state department for a further five years. Atkinson himself is a chemist on secondment from the University of Arizona at Tucson.

Atkinson took up his position on

23 September, succeeding Norman Neureiter, a chemist and former manager at Texas Instruments. The post was created three years ago by then secretary of state Madeleine Albright, in response to a 1998 report from the National Academy of Sciences. The report noted that science and technology were "not receiving adequate attention within the department" (see *Nature* 395, 313; 1998).

Despite a small increase since then in the number of scientists working with the department, Atkinson believes that lack of scientific expertise is still a "potentially dangerous situation". Many scientific issues, from carbon emissions to visa rulings for incoming researchers, fall in its remit.

The new funding will support 15 fellows over the next three years, Atkinson says. Besides increasing fellowship numbers, he



George Atkinson says a lack of science expertise is potentially dangerous.

also hopes to devise more ways for researchers to act as consultants to the department when their fellowships are over. His office also plans to hold quarterly briefings for senior department officials on the latest developments in science and technology.

Atkinson faces a

busy time both at home and abroad, says Irving Lerch, head of international affairs at the American Physical Society. As well as dealing with domestic issues such as delays to visiting researchers' visa applications, Atkinson is expected to help rebuild Iraq's scientific infrastructure. The country's fledgling science agency has so far had little interaction with US scientists, says Lerch. "I think that George is faced with a serious problem there." ■

♦ www.state.gov/g/stas

M. LYNN/UNIV. ARIZONA

Research council plan for Europe gets up steam

Alison Abbott, Munich

Europe needs an independent research council to support basic science — and the European Union (EU) should pay for it, a high-level group of experts has concluded.

The expert group, commissioned last year during the Danish presidency of the EU, has published an interim report outlining options for the creation of a continent-wide European Research Council (ERC).

The report will now form the basis of discussions with interested parties, including governments and scientists (see also page 451). The group's final report will be delivered to the Danish research councils in mid-November.

The debate about the ERC has been simmering for years. Almost everyone agrees that a transnational agency is needed to promote excellence in basic research in a continent where funding for science is fragmented along national lines. And most experts think

that the Framework programme, Europe's existing, heavily bureaucratized research scheme, is not up to the task, partly because of its emphasis on applied research.

But arguments over who should provide the funds have dominated discussions. Labouring the old linguists' joke that the first phrase to learn in a foreign language is "my friend will pay", some experts have argued that the EU should stump up the bulk of the cash, whereas others counter that member states should foot the bill, probably through their existing national agencies.

The expert group — chaired by Federico Mayor, a molecular biologist from the Autonomous University of Madrid and a former director general of the United Nations Educational, Scientific and Cultural Organization — falls squarely into the former camp. "We took as our starting point recent declarations made by EU heads of state that Europe cannot be competitive in the knowledge-based economy without a significant increase in research funding," says Mogens Flensted-Jensen, vice-chairman of the expert group and a representative of the Danish Research Councils. Flensted-Jensen also dismisses the idea that the ERC could be founded by adapting an existing organization, such as the Strasbourg-based European Science Foundation.

"If the ERC had to rely on making small improvements on the workings of current instruments like those of the European Science Foundation and the EU Framework programmes, and a bit of extra money from here and there, we are wasting our time," Flensted-Jensen argues. "Our group was very clear on the point that the initiative needs new money."



European research: who should pay?

The report recommends that a specific line should be created for the ERC in the EU's spending plans, which should be approved by the European Parliament. It suggests an annual budget of E2 billion (US\$2.3 billion).

The document also says that the ERC should be politically accountable to the European Parliament and its member states. Funding priorities and policies — such as mobility for researchers and access to large research facilities — should be agreed in an open manner with political bodies, it says. "But then the final distribution of grants should be decided on scientific quality without further political or geopolitical influence," says Flensted-Jensen.

European research commissioner Philippe Busquin says that the commission will set out its own report by the end of this year on how an ERC could be designed and funded. ■



Wellcome to fund publication in open-access journals

Declan Butler, Paris

One of the world's largest research charities, the UK-based Wellcome Trust, has lent its support to calls for 'open access' to the scientific literature.

A report to be released by Wellcome this week — *An Economic Analysis of Scientific Research Publishing* — says that the current system of thousands of subscription journals "does not operate in the interests of scientists and the public, but is instead dominated by a commercial market intent on improving its market position". In the report, the trust announces that it will allow scientists that it funds to use their grants to pay author charges required by open-access journals.

In a separate statement, the trust also pledged its support for online journals, such as those of the Public Library of Science

(PLoS), which will test alternatives to the 'reader pays' model of most research journals, which charge readers or libraries for subscriptions. Such open-access journals aim to transfer all publishing costs up front as a 'dissemination' fee paid by authors or their institutions, with papers then being made available free online.

"As a funder of research, we are committed to ensuring that the results of the science we fund are disseminated widely and are freely available to all," says Mark Walport, director of the Wellcome Trust. "Unfortunately, the distribution strategies currently used by publishers prevent this."

One of the obstacles to adoption of the alternative models may be scientists' reluctance or inability to pay dissemination fees — often around \$1,500 — when they

can publish for free in subscription journals. By endorsing the use of its research grants to pay such fees, the trust is supporting the idea that publication costs should generally be included as part of overheads on research spending.

The US-based Howard Hughes Medical Institute has already agreed to provide its investigators with up to \$3,000 each in 2004 to cover open-access dissemination fees. In its report, Wellcome calls on other funding bodies to adopt similar policies. "The fundamental point is that as a research funder we have to question whether it is right that we, and others, are in the position of having to pay to read the results of the research that we fund," says Walport. ■

Next week's *Nature* will include a News Feature analysing business models for open-access publishing.

China takes centre stage for liver proteome

David Cyranoski, Changsha, China

China is set to lead a massive research project to describe all of the proteins in the human liver — the liver proteome. The initiative is being coordinated by the Human Proteome Organisation (HUPO), an international group that is also overseeing plasma and brain proteome projects.

But parts of China's plan are controversial — in particular, a project to generate a range of different antibodies in one go.

Last month, China surprised the world of protein chemistry by pledging 200 million yuan (\$24 million) for the three-year pilot phase of the international liver-proteome study. Two weeks later, researchers at China's annual proteomics meeting, on 18–21 September, established a Chinese branch of HUPO and elected as its director the ambitious Fuchu He, a systems-biology researcher at the Beijing Institute of Radiation Medicine. "This is a golden opportunity for China to lead an international effort," says He, who will coordinate activities in up to 20 countries including Canada, France and the United States.

He says that China is drawn to the project because of the toll that liver diseases such as hepatitis and hepatoma take on the population. According to the World Health Organization, China has one-third of the world's carriers of hepatitis B, and each year the disease kills 280,000 people there.

The project aims to identify all of the tens, or hundreds, of thousands of proteins that are expressed in the human liver. It will also collect information about their concentrations in normal and diseased states, where they are localized within the liver, which other proteins they interact with, and whether they exist in various chemically modified forms.

The pilot phase has already begun. French scientists will coordinate sample collection, which will gather normal and diseased tissue taken during transplant surgery from African, Asian (Mongolian) and Caucasian populations. By the end of 2005, the project's planners hope to have data on some 3,000–5,000 liver proteins. By 2010 they hope to have pinned down 10,000 proteins.

One of the project's most valuable resources could be a huge collection of monoclonal antibodies that a team at the Beijing Institute of Radiation Medicine is planning to generate and make available at cost to scientists around the world.

The team, led by immunologist Qihong Sun, intends to make antibodies against 5,000 proteins by the end of 2005, using a technique that they consider to be particularly suitable for large-scale analysis. Con-



Fuchu He (left) has the task of coordinating China's project to identify liver proteins.



ventionally, antibodies are made one at a time, but Sun's team will inject several different proteins extracted from human livers into mice and then harvest the antibodies made to each of them from the mice's blood. These antibodies will be identified individually by identifying the specific protein to which each binds. "This is a fast approach, particularly for the highly abundant proteins we will be analysing first," says Sun.

But the antibody project might prove to be a diplomatic headache for HUPO, which is trying to standardize practices among the various proteome projects. "The dust has not settled yet on the choice of strategy for antibody production," says HUPO's president Samir Hanash of the University of Michigan.

Helmut Meyer of the University of Bochum, Germany, who heads HUPO's brain-proteome project, thinks that the Chinese researchers may be biting off more than they can chew. "Sometimes it's nearly impossible to identify the protein to which an antibody binds," he says. His brain-proteome project plans to start an antibody project next year using single purified or synthesized proteins to generate antibodies.

This issue will be thrashed out at HUPO's second international meeting in Montreal, Canada, next month. According to He, this meeting will also address the issue of continued funding for the liver proteome project. The Chinese have made a tentative promise to invest a further 1.8 billion yuan to carry the project through to 2010, but he hopes that other countries will also contribute. ■

Planck makes permanent posts

Quirin Schiermeier, Munich

In some ways, the cure became part of the disease. Rules set up in Germany last year, intended to improve career paths for researchers, restricted scientists to a maximum of 12 years of public funding.

But young scientists who stay in Germany may need this time just to complete their PhD, one postdoc and one contract as an independent junior research-group leader in a university or research institute — even though a permanent academic job such as a professorship usually requires more experience than this.

Now, in a surprise move, the Max Planck Society (MPS) will relax its strict non-tenure recruitment policy in a bid to help some of the heads of junior research groups in its 80 research institutes to stay in the system.

The society's rules allow only one five-year contract as a junior research group

leader. But the MPS now plans to allow the institutes to offer its best junior research heads *de facto* permanent positions when their five-year contracts run out.

"We need to give young scientists more time to develop the experience and track record they require," says Herbert Jäckle, a director at the Max Planck Institute of Biophysical Chemistry in Göttingen and vice-president of the MPS. "Otherwise, most of them will be forced, against their will, to continue their careers abroad, or outside the academic system."

Jäckle says that only the potential high-flyers among the society's 100 or so young group leaders will be offered extensions. The MPS likes researchers to leave when they have a strong publication list, to take up jobs at universities. "It is of course possible that some might stay forever," says Jäckle, adding that "this is a risk we're ready to accept." ■

Minister knocked down by court verdict on mosque demolition

New Delhi India's science minister, Murli Manohar Joshi, has submitted his resignation in the wake of a scandal over the destruction of an ancient mosque more than ten years ago.

In December 1992, fanatical Hindus tore down the Babri Masjid mosque at Ayodhya in the state of Uttar Pradesh, as they wanted to rebuild a Hindu temple they believe once stood on the same site. The act sparked nationwide rioting and fighting between Hindus and Muslims that resulted in hundreds of deaths.

On 18 September this year, an Uttar Pradesh court upheld allegations that Joshi, along with six others, instigated the mosque's destruction. The court has ordered that the group face criminal charges, although Joshi maintains his innocence and has filed a petition against the ruling. India's deputy prime minister, Lal Krishna Advani, was also named in the case but has been exonerated.

As *Nature* went to press, Joshi's resignation had not been accepted by India's prime minister, Atal Behari Vajpayee.



Demolition squad: the Babri Masjid mosque was destroyed by Hindu fundamentalists in 1992.

Money runs out for project to spy on bank transactions

Washington The US Congress has axed a Pentagon intelligence office that has come under fire from civil-rights advocates.

The Information Awareness Office was part of the Defense Advanced Research Projects Agency (DARPA), an arm of the defence department. The office was criticized last year for a plan to comb the financial details of millions of citizens in search of potential terrorists.

The project was the brainchild of former DARPA director John Poindexter, who resigned last month amid a row over another project, in which participants could bet on future terrorist attacks (see *Nature* 424, 601; 2003).

Many US security experts argued that the projects should not have been terminated, despite protests from civil-rights groups that they were unethical.

Venter's dog's genome sets tongues wagging

Washington First he sequenced his own DNA. Now Craig Venter has sequenced the DNA of his pet poodle, Shadow (pictured).

Choosing to sequence a dog is a wise move, as there are more than 350 known canine genetic diseases — more than for any other animal apart from humans — although picking a poodle may seem a curious choice. A US government project, expected to be completed early next year, has chosen to sequence a boxer, as this is one of the least genetically variable breeds and is likely to give a representative sequence.

But dog breeds are more than 99% identical to one another, so Shadow's sequence should be adequate for studying some diseases. It should also help track canine evolution. Based on the sequence, which is 80% complete, it is estimated that 18,473 dog genes have human



equivalents — slightly more than are known to be shared by humans and mice (E. F. Kirkness *et al. Science* 301, 1898–1903; 2003).

California rejects party line on stem-cell work

San Diego More human embryos could be available for research in California as of January 2004, thanks to legislation signed into law last week.

The law requires staff at fertility clinics to ask all patients if they are happy for their discarded embryos to be used for research. Currently, the law only suggests discussing the option of embryo donation with patients.

This ruling is part of a raft of legislation regarding stem-cell work in California — in spite of a federal decision in 2001 to limit such research. A registry of donated embryos is planned, and a 13-member commission of scientists, ethicists, attorneys and religious leaders will be in place by 2005 to provide guidelines for the work. Some state legislators are also pushing for hundreds of millions of dollars in bonds to create a pool of stem-cell research funding.

Brazilian ruling sows seeds for transgenic future

Washington Brazil, one of the world's largest exporters of soy beans, has legalized the planting of transgenic versions of the crop for the coming growing season. The government's decision is expected to pave the way for the permanent legalization of genetically modified crops.

Brazil's courts banned transgenic crops in 2000, but farmers in the south of the country have long been suspected of importing transgenic seeds illegally from Argentina. Currently, 20–30% of Brazil's soybean crop is thought to contain transgenes. Last year the government granted an exemption from the ban, initially for a single growing season. That concession has now been extended for another year.

Brazil's official 'transgenic-free' status has allowed it to continue selling produce to European countries that refuse to import transgenic food. But that could soon change.

Nigeria joins the space age with first satellite

London Africa took a big step into space on 27 September with the successful launch of a Nigerian Earth-monitoring satellite.

The craft, NigeriaSat-1, will monitor Nigerian territory to help prevent disasters caused by floods, landslides, oil spillages and fires, or to assist relief workers during such emergencies. It will be operated by the country's National Space Agency in Abuja.

The satellite will form part of a disaster-monitoring constellation — a cluster of satellites built by the British company Surrey Satellite Technologies. An Algerian contribution to the cluster was launched last November, and two other satellites — from Turkey and Britain — were launched with NigeriaSat-1 by a Russian booster. A Chinese satellite will join them in early 2005.

NigeriaSat-1 is the first product of a €20-million (US\$23-million) investment in space science made by the Nigerian government in 2001.

Correction

Due to an editing error, last week's article "SARS triggers biomedical shake-up in China" (D. Cyranoski *Nature* 425, 333; 2003) incorrectly states that most of China's biomedical research is supported by the Chinese Academy of Sciences when in fact the responsibility is divided among several governmental bodies. The article also incorrectly implies that the academy aims to retain this central position in biomedical funding by resisting new initiatives. In fact, many leaders of the academy strongly support the initiative described in the article and have been instrumental in introducing it to China's prime minister.



Across the great divide

Scientists like to think that research collaboration can overcome political barriers. But for those on opposite sides of the Israeli–Palestinian conflict, how realistic is this ideal? Jim Giles visited the region to find out.

There can't be many coffee-break conversations as tense as some of those that take place in Gershon Golomb's lab at the Hebrew University of Jerusalem. Golomb, who heads a research group working on improved methods of drug delivery, lives in Efrat, an Israeli settlement in the occupied territory of the West Bank. To Palestinians, the very existence of such towns is a major barrier to peace. Golomb is also a senior officer in the Israeli reserve forces, and until recently served in the army for up to two months every year, often on the West Bank. All of this makes it surprising — and unique, claims Golomb — that one of his team is a Palestinian.

That researcher, Yusef Najajreh, lives in Beit Jala, just a few kilometres from Efrat. Even nearer to Najajreh's hometown lies the Israeli settlement of Gilo — to which Irith Gati, the group's technician, returns home each evening. When violence flares overnight — as it frequently does — the group's discussions the next morning are understandably strained. "There can be shooting between Gilo and Beit Jala," says

Najajreh, whose house has been hit by stray bullets. "I come in and we accuse each other's community of starting it. But at the end of the day we are there for the science."

To those who believe that scientific collaboration can help break down the barriers raised by conflict, the fact that Najajreh and his Israeli colleagues are able to work together against the common enemy of disease is an encouraging sign. And although few working relationships are overshadowed quite so starkly by the conflict as those in Golomb's lab, examples of joint Israeli–Palestinian research projects are not hard to find — a paper in this issue of *Nature*, for instance, on the seismology of the Dead Sea region, boasts both Israeli and Palestinian contributors (see page 497).

But it would be naive to conclude that such collaborations tell a simple story of scientists successfully putting their political differences aside in the pursuit of knowledge. Researchers from both communities are able, in many cases, to manage their huge divergences in viewpoint — but these differences cannot simply be disregarded. Would

you expect an Israeli academic who has lost a friend in a suicide bombing to have no misgiving about working alongside people from a society that sees the perpetrator as a martyr? And is it any surprise to learn that Palestinian researchers who are shut out of their labs by the Israeli army, and whose neighbours' homes have been bulldozed, may view the idea of collaborating with Israeli scientists with incredulity?

For an outsider, even finding the right language to discuss the situation is problematic. When e-mailing scientists to set up meetings before my visit to the region, I inadvertently offend one Palestinian researcher by asking about "science on the West Bank". He requests, politely but firmly, that I refer to the territories occupied by Israel as "Palestine". But for many Israelis, his preferred label for this disputed land is equally inflammatory.

In the end, Israeli and Palestinian scientists alike ignore my clumsy attempts to find acceptable terminology. They are happy to discuss their work, and how it has been affected by the conflict — or, in rare cases,

defined by it (see 'Know the enemy', overleaf). I visit in August, a period of relative calm. Hamas and other militant Palestinian groups have called a ceasefire, and the Israeli army's presence in the occupied territories is more low-key than in the preceding months. As a result, I am able to venture onto the West Bank to meet Palestinian scientists without encountering too many problems. In Israel, students are on their summer break and there is a relaxed atmosphere on the modern, Western-style campuses.

Yet within two days of my departure, the violence and despair have returned. Today, the outlook is as bleak as anyone can remember. Suicide bombings have resumed, bringing terror to the heart of Israeli society, and Israel's military has embarked on a vigorous security clampdown, leading to the deaths of around 40 Palestinians. One Israeli minister has even aired the possibility of assassinating Yasser Arafat, president of the Palestinian National Authority.

Such hardline attitudes find few echoes among Israel's academics. Like their counterparts in many other countries, their political views tend mainly towards the liberal end of the spectrum. Many researchers criticize their government's strategy in the occupied territories, and argue that politicians should do more to push the peace process forward. What's more, Israeli academics are generally enthusiastic about the idea of



working with Palestinian scientists. "A lot of Israelis hate the idea of sitting in an ivory tower and doing nothing," says Benjamin Geiger, a cell biologist at the Weizmann Institute of Science in Rehovot. "Over the past few years there have been many, many attempts to create interactions with Palestinian colleagues."

Some of these attempts have failed because of the huge disparity in resources at

the disposal of these neighbouring research communities. Israel is a world leader in fields such as biotechnology and physics, and the country's labs are as well-equipped as those anywhere in the world. By contrast, the total amount spent on research across all of the Palestinian universities is a fraction of that deployed by a single major Israeli research institution. Sometimes, Israeli scientists who are interested in reaching out to Palestinian colleagues simply fail to find a partner.

But numerous collaborations do exist, spanning fields from chemistry to plant biology. The Israeli researchers involved stress the scientific value of these efforts, but many are also motivated by a desire to promote peace by helping their Palestinian colleagues to build up their research capacity. "We know that building science communities is important," says microbiologist Hervé Bercovier, vice-president for research at the Hebrew University, which has been particularly active in promoting joint projects with Palestinian institutes.

In some fields, the results of collaborative projects will feed directly into any peace negotiations. Some of the region's key aquifers lie beneath the West Bank, and hydrologists frequently find themselves involved in political arguments about how these resources should be distributed (see 'Water and the wall', below). "If we want to share water in a professional manner, we

Water and the wall

When it comes to juggling science and politics, the hydrologists of the Middle East are experts. In this arid environment, deciding how to allocate the freshwater of the Jordan River and the aquifers below the West Bank will be a vital component of any peace agreement.

"Water in Palestine is not hydrogen and oxygen," says Abdel Rahman Tamimi, director of the Palestinian Hydrology Group, based in Ramallah. "It's politics." The sensitivity of the issue is highlighted by the fact that data on the amount of water extracted from aquifers by Israeli settlements on the West Bank have been declared a military secret.

Yet despite such obstacles to cooperation, Israeli and Palestinian hydrologists have a surprisingly good record of working together — producing reports that may prove invaluable in future political negotiations. Funding from abroad, meanwhile, has enabled new monitoring stations to be set up, improving knowledge of seasonal river flows and sources of pollution.

Today, however, there is a new problem: the 'security wall', now partly constructed, that will separate the West Bank from Israel. Palestinian hydrologists claim that it is designed partly to annex key water resources. The barrier, which the Israeli government says is necessary to restrict the movement of terrorists, at some points snakes several kilometres beyond the pre-



High and dry: Palestinians have accused Israel of using the security wall to annex water supplies.

1967 Israeli border, and has left some Palestinian farmers unable to reach their land. On the edge of Jerusalem, it slices through the campus of Al-Quds University, forcing staff and students to take detours lasting several hours to get to work.

The wall has also separated some Palestinians from their water supplies. When a World Bank-led team visited the initial section that had been built on the northwest border of the West Bank in May, it found a handful of sites where the barrier comes between Palestinian communities and the wells that they had used for irrigating their crops.

The Israeli authorities say that the wall's path is determined by the need to protect vulnerable Israeli settlements. They argue that problems with

access to water will be resolved by putting gates in the wall, and by issuing permits to cross it.

But Palestinian hydrologists reject this explanation, and are convinced that the appropriation of water resources is one of the primary goals behind the wall's construction. "This will finalize the status of water rights before negotiations begin," Tamimi complains.

So far, only around 150 kilometres of wall have been built, and the Israeli authorities have not released details of its future direction. But by talking to villagers who say they have been approached by the Israeli army about requisitioning their land, Tamimi's hydrology group and other Palestinian non-governmental organizations have drawn a map of the course they believe the wall will take. This suggests that it will leave some 50 wells used by Palestinians on the Israeli side.

Israeli hydrologists see things differently. "The basic reason for the wall is security," argues Hillel Shovel, professor emeritus at the Hebrew University of Jerusalem and one of Israel's most prominent hydrologists. "I realize that injustices are being done, and these need to be corrected. But I don't accept that the goal is to steal water resources. That is paranoia."

Despite the proud record of cooperation between Israeli and Palestinian hydrologists, it seems some subjects are just too politically sensitive for the two sides to reach agreement.

need to work together," says Gedeon Dagan, a geophysicist at Tel Aviv University who sits on the steering committee for a project on the hydrology of the Jordan Valley involving both Israeli and Palestinian scientists.

Talk to Palestinian researchers about the value of joint projects with Israeli scientists, however, and you hear a diverse set of opinions. The Palestinian Ministry of Higher Education opposes links with Israeli institutions. And although many Palestinian researchers ignore this official line, many of those who work in towns that have borne the brunt of security operations by the Israeli army support their government's view.

At An-Najah National University in the West Bank town of Nablus, chemist Maher An-Natsheh explains why his experience of the conflict has left him unwilling to work with Israeli scientists. "We are in a state of war," he says. "They have to give Palestinians the right to exist; then we can start talking about collaborations."

An-Najah University's students often vote for militant groups such as Hamas in campus elections, and the university is regarded as a nest of terrorists by the Israeli authorities — the army says that several suicide bombers have been An-Najah students.

Two months before my visit, Fadi Alawneh, a journalism student at An-Najah, tried to avoid Israeli army checkpoints by crossing a deep ditch; he lost his balance when approached by an armoured vehicle and died from his injuries. He was the thirty-fifth An-Najah student to die since the present Palestinian uprising, or *intifada*, began in September 2000, says An-Natsheh.

During this period, An-Najah University has been closed several times. An-Natsheh, who is the university's vice-president for academic affairs, says that deliveries of chemicals are frequently impounded on suspicion that they may be used to make explosives. In



Israeli tanks are once again a common sight on the streets of West Bank towns such as Nablus.

Hebron, some 80 kilometres away, where tensions are particularly high due to the presence of a small Israeli settlement surrounded by a hostile Palestinian community, it is a similar story (see 'A campus under siege', page 449).

West Bank towns are also frequently held under curfew. Research at many institutions ceases during such clampdowns, but some scientists take risks to continue their work. In July 2002, for instance, a curfew order prevented staff at the Applied Research Institute-Jerusalem, which conducts environmental studies, from commuting to and from their headquarters in Bethlehem. Rather than stop working, they moved computers and equipment to the houses of friends and family in Beit Jala, where the narrow streets made it easier to evade the Israeli

army. "You could easily knock on a door and hide if you saw a patrol," says Jad Isaac, the institute's general director.

Even against this background, some Palestinians are open to the idea of interacting with Israeli scientists. But they are not always comfortable about expressing this view in public, for fear of attracting a backlash from their own community. Last month, for example, a small group of Palestinians attended a 'Frontiers of Science' conference in Istanbul, Turkey, which was organized by the US National Academy of Sciences with the aim of promoting dialogue between researchers from across the Middle East. But some of these researchers were anxious about the signals that they were sending out by attending. "We don't want the media to take pictures and announce that Palestinians met Israelis," says Awni Khatib, a chemist at Hebron University.

Other Palestinian researchers are simply so frustrated with the difficulties of daily life that they have no wish to work with scientists from the country they see as the cause of their troubles. Even if researchers from Nablus and Hebron did want to collaborate with Israeli scientists, they would struggle to travel to the nearest Israeli universities in Jerusalem, just a few dozen kilometres away. The journey would require a special permit, which can take months to arrive and still does not guarantee access.

Travelling back to Jerusalem from Hebron one afternoon, the reality of this situation is brought home. The taxi-van in which I'm travelling is stopped at a checkpoint and instructed to turn back — the road has been closed for security reasons. Nearing the end of a hot and uncomfortable journey, our driver has other ideas. He initially withdraws as requested, but quickly cuts back onto the main road and speeds past the

Know the enemy

In May 2000, Ariel Merari persuaded an extraordinary group of people to sit around the same table. Several armed organizations, responsible for kidnappings, killings and bombings, sent representatives to Paris. There, away from the world's media, a group of academics probed the attitudes of people who have been vilified by their opponents as heartless terrorists.

In what Merari describes as a "cosy and informal setting", political scientists and psychologists met with members of groups including the Basque separatist organization ETA, the left-wing Colombian guerrilla group FARC, and armed militants associated with Yasser Arafat's Fatah faction of the Palestine Liberation Organization. Through various role-plays, the militants were asked to act out the side of governments and armed rebels involved in

confrontations.

Merari has made a career out of trying to understand people involved in armed uprisings. Trained as a psychologist at the University of California, Berkeley, his career trajectory changed dramatically after the 1973 Yom Kippur War. Then a reserve paratrooper, Merari was sent to the Golan Heights. His unit was trying to rescue colleagues pinned down by Syrian fire. "We were running towards the Syrians when I got a bullet in the chest," says Merari. "I was evacuated. If I had got to hospital two minutes later, I wouldn't be sitting talking to you now."



Ariel Merari: seeking to discover what motivates armed militants.

After this experience, Merari felt incapable of returning to his research on hormones and behaviour. "This ivory tower suddenly seemed too detached from the reality of Israel," he says. In the mid-1970s, his expertise as a psychologist earned him a place on an army hostage-negotiation team. Around the same time, he began to trawl the literature for work on terrorism — a direction that ultimately led him to Tel Aviv University to study the subject full-time.

Merari began to focus on suicide bombers in the early 1980s. One of his methods is the psychological autopsy. Through a trusted third

checkpoint, escaping the young soldier's attention. "Don't worry, he would have shot our tyres before aiming at us," another passenger says.

Not all Palestinians face such severe restrictions on their movements, however. And where security controls are less stringent, attitudes towards working with Israelis are much more positive. At Al-Quds University in Abu-Dis, a suburb of eastern Jerusalem, Palestinian researchers are building links with nearby Israeli universities, with encouraging results. "Before we started collaborating in 1994, Al-Quds spent US\$35,000 a year on research," says Ziad Abdeen, who works on nutrition and disease and is the university's dean of research and graduate studies. "Now it is \$3 million."

Most of this money has come from abroad, and much of it was made available specifically to promote joint projects with Israeli institutions. The increased funding, says Abdeen, has provided new facilities such as the university's molecular-biology lab, constructed using money from the Belgian government for a joint project with the Hebrew University. Israeli academics have also helped researchers at Al-Quds to gain experience in writing grant proposals. "Our research culture was created by working with our Israeli colleagues," says Abdeen.

The issue of collaborative projects is not the only one to split the Palestinian research community. In April 2002, more than 100 academics, mostly Europeans, wrote to the British newspaper *The Guardian*, calling for the European Union to suspend Israel's participation in its Framework research programme, in protest at "the violent repression of the Palestinian people". The idea of boycotting Israeli science was rebuffed by the EU, but generated a remarkable amount of debate. Isolated incidents of academics refusing to work with Israeli colleagues were

Show of support: Palestinian women wave banners bearing the insignia of the militant group Hamas.

party, Merari has conducted interviews with friends and families of most of the 36 *shaheeds* — the Arab word for martyr — who struck Israeli targets up until 1998. By combining these interviews with studies of bombers who were apprehended before they detonated their devices, he has been able to build up psychological profiles of the attackers.

None of the bombers seemed to need psychiatric help, nor were they feeling suicidal in the normal sense of the word. They came from a broad cross-section of Palestinian society. "There was no single psychological profile," says Merari.

More important than individual characteristics, Merari claims, is the role of the group that helps to organize the bombings. Many suicide bombers, Merari believes, volunteer in the heat of the moment, and are then placed under a 'contract of honour' from which they may find it

difficult to back out. "Shortly before they are sent on their mission, most *shaheeds* are filmed declaring that they wish to be a martyr," he says. "It's hard to break something like that."

Equally important, Merari argues, is the role of a society that idolizes suicide bombers — posters depicting *shaheeds* are a common sight in Palestinian towns. "In this kind of atmosphere, many people say: 'I want to become a *shaheed* too'," says Merari.

Eyad El-Sarraj, director of the Gaza Community Mental Health Programme, argues that such attitudes are borne of a desperation to be free from Israeli rule. He says that many Palestinian youths feel that they should sacrifice their lives for the good of their people — a view that is strengthened by the belief that they will be rewarded for their actions in heaven.

In the long run, Merari believes that countering

the wave of suicide bombings will require a sea change in Palestinian public opinion. He suggests that the Israeli army might become involved in distributing food and medical care in Palestinian towns, so that residents' experiences of the security forces are less uniformly negative.

For many Israelis, however, Merari's ideas are unpalatable. Following the recent upsurge in violence, security clampdowns are the order of the day. And for most Palestinians, Merari's suggestions simply miss the point — they want the Israeli army to withdraw from their towns, not to act like an aid agency.

For all of his desire to make a practical contribution towards peace, Merari is forced to admit that he has yet to reach out beyond the confines of academia. Asked whether politicians listen to his arguments, he responds: "Unfortunately, very little."



Bus blast: more than a dozen Israelis died in this bombing in Jerusalem in June; after a period of relative calm, such attacks have now resumed.

reported, and the controversy was re-ignited in June this year, when a pathologist at the University of Oxford, UK, was suspended from his post after rejecting an application from a prospective PhD student who had served in the Israeli army (see *Nature* **424**, 120; 2003).

Many Palestinian researchers reject the idea of boycotting Israeli science. "It's counter-productive," says Abdeen. Even among Palestinians who support the idea

of an economic boycott on Israel, there is unease about the idea of extending protests into the scientific arena. In towns such as Nablus and Hebron, however, some researchers support the actions of those foreign academics who have refused to work with Israeli scientists. Despite his own willingness to work with Israelis if the collaboration strengthens Palestinian science, Khatib is among their number. "Pressure must be exerted on Israel," he says. "I support the

boycott because it can enhance movements in this direction."

Such comments are distressing to Israeli academics, especially those involved in collaborations with Palestinians. There is a saying in Israel that reflects the country's love of debate: "put two Israelis in a room, and you'll get three opinions". But this definitely does not apply to discussions about the call for a scientific boycott. "You're not going to get three opinions on that," says Jonathan

Touched by terror

For almost a year, the unpainted door of Michael Beenstock's office in the Department of Economics at the Hebrew University of Jerusalem served as a terrible reminder of the darkest day in Israel's academic history.

In July last year, builders were busy renovating Beenstock's department and other buildings on the university's Mount Scopus campus. On the final day of that month, one of the workers left a bomb in a cafeteria, close to Beenstock's office. Nine people — staff, students and visitors — were killed, and more than 80 injured. "Every day, that unfinished door reminded me of what happened," says Beenstock.

For academics who believe that research collaboration can help to defuse Israeli-Palestinian tension, the targeting of the Hebrew University was especially painful. Of all the universities in Israel, it has the strongest tradition of organizing joint projects. More than 20 collaborations with Palestinian researchers are still ongoing, and 45 new Palestinian students have enrolled for courses this autumn.

Memories of the bombing are still vivid.

Beenstock was in his office at the time, talking with some of his students. "We heard a bang and thought the builders had caused an accident," he says. "Then we heard ambulances and knew it was a bomb."

Yousef Najajreh, a Palestinian researcher at the university who works on methods for drug delivery, was having lunch in a cafeteria on another of the university's campuses when the bomb went off. "My wife didn't realize there were different cafeterias," he says. "She knew I would be eating at that time. She was going crazy trying to call me, but the telephone network was down."

Najajreh condemns the attack on an institution that has tried to build links between Israelis and Palestinians, but says that he nevertheless found it difficult to approach Israeli colleagues in its immediate aftermath. "It's not easy for a Jew to see a Palestinian after people have been killed," he says.

Beenstock remains angry about the muted reaction from many of the Hebrew University's Palestinian students. On the day of the bombing, he had spoken to an Arab postgraduate whose

PhD thesis he had supervised. "People from all over the world called that night," recalls Beenstock. "But my student never bothered to pick up the phone."

Clearly, the bombing has scarred relationships between Israelis and Palestinians at the university. But officials remain determined to continue the tradition of joint projects. Hervé Bercovier, the university's vice-president for research, notes that a contract to collaborate with An-Najah National University in Nablus — which has been accused by the Israeli army of harbouring terrorists — was signed earlier this year. "There are no restrictions on who we work with," he says.

That sentiment is shared by some of those most severely affected by the bombing. Inna Zusman, then a first-year Israeli student in computing and cognitive science, suffered a spinal-cord injury in the blast. Now in a wheelchair, she will resume her studies this autumn. "People look more at your face and your race than before," she admits. "But the attack won't change the way we interact."

Gressel, a plant biologist at the Weizmann institute. "I don't think it ever helps to keep scientists out of science for political reasons." His colleague Geiger agrees: "This breaks with one of the most cherished and important features of science: that it is international and non-political."

Scientific bodies and publications, including *Nature* (417, 1; 2002), have spoken out against the idea of boycotting Israeli science, and an anti-boycott petition has attracted more than 15,000 signatures. The academics behind the letter to *The Guardian* have also since stressed that any boycott should not extend to Israelis who work on joint projects with Palestinian researchers.

Nevertheless, many Israeli scientists feel that the debate has had a chilling effect on their relations with colleagues abroad. When I suggest to Shy Arkin, a biochemist at the Hebrew University, that the boycott is only weakly supported in Britain, he immediately dissents: "It's not a small minority."

"You hear of a variety of initiatives," agrees Geiger. "Some are individual people expressing their views; others are channelled in a more institutional way." Geiger points out that the European Molecular Biology Organization has come under pressure from some academics to suspend the membership of Israeli scientists.

Other Israeli academics tell of more personal attacks. "There is an unpleasant feeling in Europe," says Michael Beenstock, an economist at the Hebrew University. Just back from a meeting in Finland, Beenstock recounts an argument with an academic who began to berate him about his government's policies towards the Palestinians, after reading the institutional affiliation on his name badge. "He very quickly became critical of 'you Israelis'," says Beenstock. "I lost my temper until someone told us both to shut up."

It is understandable that Beenstock might lose his cool under such circumstances — his office is less than 100 metres from the site of last summer's attack on a Hebrew University cafeteria, in which nine people were killed by a bomb placed by Palestinian construction workers (see 'Touched by terror,' opposite). For those who have been exposed to such acts of violence, being blamed for the conflict is hard to bear.

Back in their drug-delivery research lab, Golomb and Najjreh try not to engage in similar arguments, as they know that they can never agree over the conflict. Yet somehow they manage to work together. "We turn down the volume on our arguments because we have a shared value against sickness," says Najjreh.

Until the political leaders of Golomb's and Najjreh's respective communities can negotiate a peace agreement that both sides can accept, however, this shared value will remain bound by a tenuous thread. ■

Jim Giles is a reporter for *Nature*, based in London.



Up in smoke: in recent weeks, the Israeli military has targeted buildings suspected of housing Palestinian militants in Hebron.

A campus under siege

Radwan Barakat waves his hand towards the olive groves and low-rise buildings that blanket the hills opposite Hebron University. "They call this the martyrs' neighbourhood," he says. The area's association with suicide bombings — or 'martyrdom operations', in the lexicon of Hamas and other militant Palestinian groups — has left its mark. Across the valley, a single apartment has been blown out of the side of a building. This is the usual response when the Israeli army identifies the home of a suicide bomber.

The dusty campus of Hebron University, which hosts some 4,500 students, also bears the scars of conflict. Barakat takes me in through a side entrance — the only one that is not welded shut. The university was shut down in January, after students from the area were linked with suicide bombings. This single gate was broken open a few months later.

During the closure order, which lasted until shortly after my visit, the university's activities were in disarray. Barakat is a plant pathologist, and needed to care for his plants and fungi to avoid losing data. He points to a low roof below the windows of his laboratory, which allowed his support staff to enter without attracting the attention of soldiers who were watching the building's main entrance. "I really have to respect my technicians and graduate students for doing this," Barakat says.

Other areas of Barakat's work have not survived, however. He used to divide his time

between the university and an agricultural research station at Al-Aroub, a few kilometres outside Hebron. But this station sits opposite a refugee camp that is closely monitored by the Israeli army. When the latest Palestinian uprising began in September 2000, Israeli forces moved into the station. A watchtower now stands where Barakat's experimental crops once grew, and access to the area is prohibited.

Work at Hebron University continued despite the official closure. Nearby school halls were recruited for lectures, and students moved in as the schoolchildren left each afternoon. Teaching hours were cut back, however, to allow students to travel home in daylight hours and so avoid problems with the Israeli security forces.

On the day of my visit to the university, administrators are rehearsing for a graduation ceremony, laying out chairs and setting up a public-address system. But the next day, the Israeli army intervenes, making it clear that the ceremony should not go ahead.

When I call in September to see if the event would be rescheduled, the outlook is grim. "Listen," says Naim Daour, Hebron University's director of public relations. "Can you hear the sirens? The army has destroyed a building near the university." Later, I speak to Maher Al-Jabari, a chemist and trustee of the university. "The purpose of the celebration was to add happiness," he says. "But this now contradicts the feeling of sadness in Palestine."

NAVEF HASLAMOUN/REUTERS

Outline for a European Research Council

The ERC should start small and from scratch, using the expertise of existing bodies.

Sir — During the past two years, European scientists, politicians and science policy-makers have engaged in a great deal of discussion about the desirability of creating a pan-European funding agency for basic research, the European Research Council (ERC) (see, for example, *Nature* **421**, 881; 2003). The arguments in favour of creating an ERC are persuasive (see www.esf.org/newsrelease/63/ERC.pdf), but there remains much debate on how to organize and fund it. We propose a 'learning structure' for the ERC that would allow it to start small and from scratch, with minimal investment in capital and permanent staff, using the proven expertise of existing grant-awarding bodies.

Given the virtually unanimous opinion in favour of creating an ERC, we may optimistically assume that the stakeholders will divert sufficient funds to provide a small budget for an embryonic start-up in, say, 2004–05. The stakeholders will include the European Union (EU), national research councils (countries that currently do not have research councils should be required to establish them if they wish to participate), and perhaps also other European science-funding organizations and charities. To avoid claims of 'juste retour' — from participants wanting to ensure that they get as much out of the scheme as they have put in — countries that are far below the EU average for research funding would simply pay directly for the grants that are awarded to their own researchers through the ERC selection process rather than providing general start-up funds.

A 'learning structure' must be established to administer this start-up budget and to generate the experience necessary to build the 'grand vision' of an ERC. The ERC must have a flexible, responsive and administratively light-weight structure, learning from others' past mistakes as well as from examples of best practice. Peer review of grant applications, for example, must be rigorous and transparent, applications must be processed quickly and awards must be paid on time.

We propose that the people best placed for this role in the embryonic ERC are administrators seconded from the stakeholder funding agencies, especially national research councils. Administrators from appropriate agencies would constitute 'consortia' representing one area of research such as genomics or particle physics. Each consortium would construct a 'study section' of expert scientists to evaluate and select applications for

funding, with one administrator as coordinator. A consortium for genomics, for example, might be hosted by the European Molecular Biology Organization (EMBO), where administrators from all nations would benefit from the 'best practice' they observe at EMBO and the networking opportunities afforded by working with a multinational group. They would return to their national agencies after their secondment to the ERC having gained experience that will help to improve their country's science administration.

The administrative consortia would, however, need some central resources to create databases and websites, and the ERC would need a directorate and a headquarters — probably in Brussels, for proximity to EU organizations. The ERC will also need a president who has earned the respect of the community and who is chosen from, and approved by, a selection nominated by the national academies.

Beneath the president there would be two or three boards of governors representing broad branches of science and comprising academic leaders in the appropriate fields as well as policymakers and, where appropriate, representatives

from industry. The governors should serve a fixed term — three years, for example — with new governors being nominated and elected by a stakeholders' council made up of representatives of the funding agencies contributing to the budget of the ERC.

This interim structure is flexible and transparent, not simply administering grants, but providing a 'learning mechanism' for the national programmes and governments of less advantaged European nations to improve their domestic research funding and administration structures. In the long term, this would help to level the playing field for research across Europe.

The ideas proposed in this letter are discussed more fully in an article that will appear in the October issue of *The ELSO Gazette* at www.the-elso-gazette.org/magazines/issue16/features/features1.asp

Carol Featherstone, Kai Simons

*The ELSO Gazette, European Life Scientist Organization, 8 Rue Jean-Paul Laurens, 31450 Fourquevaux, France

†Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

How scientists can help to protect US homeland

Sir — Your news story "Research mired in Homeland Security delays" (*Nature* **424**, 986; 2003) does not reflect how the Science and Technology division of the Department of Homeland Security (DHS) is working with the private sector or the scientific community. The statement that the department is "sitting on some 3,000 unsolicited research proposals" has been corrected (*Nature* **425**, 9; 2003), but we also take issue with the overall impression given by your account.

Instead of using information provided by the Science and Technology division, your article quoted sources outside the DHS, who are unfamiliar with the operations of the department and thus unable to draw an accurate picture.

Although I understand that individuals in the scientific community are looking at the DHS as a possible source of funding, the department is not responsible for funding decisions made by other organizations, as suggested by a physicist from the Department of Energy who was quoted in your article.

The DHS is currently reviewing more

than 3,300 submissions sent in response to a Broad Agency Announcement. These initial submissions are being reviewed by the Technical Support Working Group and we expect to award contracts before the end of the year. To find more information about these proposals, interested individuals can visit our website at www.dhs.gov. Our department takes seriously its mission to coordinate the extensive talents and expertise resident in the private sector for homeland security, and this solicitation of proposals represents an important first step.

In future, I hope that *Nature* can provide a more accurate and balanced picture of the operations of the DHS to give your readers a better understanding of the role they can play in protecting the homeland.

Charles E. McQueary

Under Secretary for Science and Technology, US Department of Homeland Security, Science and Technology, Washington DC 20528, USA

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

A magical history tour

Take a trip to the end of the ice age with a fictional guide.

After the Ice: A Global Human History 20,000–5000 BC

by Steven Mithen

Weidenfeld & Nicolson: 2003. 622 pp. £25

Robert L. Bettinger

When the Pleistocene epoch gave way to the Holocene some 12,000 years ago, there was profound and rapid environmental and cultural change: the climate warmed, immense glaciers melted, rising oceans flooded ancient coasts, and an ice-age bestiary of lions, tigers, and elephants vanished. Humans around the globe were quick to exploit the opportunities created by this natural tumult and the vastly more productive environment that emerged in its wake. They developed new technologies. Some discovered and settled new continents. Others domesticated the plants and animals that feed the world today, and some of these gathered into the settled agricultural and trading communities from which the first civilizations rose.

After the Ice tells the story of these transformations. Its narrator is Steven Mithen, a well-published archaeologist who is perhaps best known for *The Prehistory of the Mind* (Thames & Hudson, 1996), a book about the evolution of human cognition. Mithen has been equally active in the field, having excavated in Scotland, and currently in southern Jordan, to shed light on the transitions that are the subject of this book. His theoretical expertise and substantial field experience are evident in the even-handed and thoughtful treatment he gives to the evidence and interpretations of the many other scientists upon whom he must rely in telling this story.

In writing *After the Ice*, Mithen set out to make prehistory more accessible to the public, but the book's scope and detail will appeal mainly to those willing to invest some time and effort in learning about this subject. There are 52 relatively short chapters, most dealing with a site or series of related sites. The chapters are collected into sections that develop a coherent picture of the Pleistocene–Holocene transition in six large geographical areas: western Asia, Europe, the Americas, greater Australia and eastern Asia, southern Asia, and Africa.

Mithen presents the archaeological and palaeoclimatic evidence in plain English, without jargon, attending to details that help the reader understand the sites, the scientists who excavated them, what they found, and the natural and cultural processes that they illuminate. The text is quite comprehensible on its own, and extensive chapter notes at the end of book provide background, explanation



Time's arrows: these 6,000-year-old cave paintings from Spain depict scenes of hunting with weapons.

and citations of original references for the serious reader. Mithen presents and evaluates all of this evidence clearly, but what distinguishes this book is the way he tries to bring it to life for the lay reader.

He creates a fictional character, John Lubbock, who has the power to travel back through time to visit the sites described by Mithen and to participate, unseen, in the daily activities of their inhabitants. Lubbock experiences what Mithen imagines it would have been like to be alive at all of these places in the past. John Lubbock is based on the Victorian archaeologist and biologist of the same name, author of *Pre-Historic Times*, *as Illustrated by Ancient Remains*, and *The Manners and Customs of Modern Savages* (1865), a copy of which the fictional Lubbock carries and consults during his travels.

The disparity between the people that the fictional Lubbock meets and the depraved hunter-gatherers depicted in *Pre-Historic Times* is the device through which Mithen seeks to correct the popular stereotype that ignorance and superstition sentenced prehistoric hunter-gatherers to a subhuman existence of drudgery and want. In keeping with the modern anthropological view, the fictional Lubbock encounters people who are largely like ourselves, only more knowledgeable about their natural surroundings and more constrained by their technology. They are intelligent, yet fearful of the unknown; industrious, yet devoted to art, story-telling and long bouts of leisure; mostly well fed, but acquainted with malnutrition and disease; and caring with family and friends, yet capable of occasional violence.

It gets a bit crowded at times, with Mithen, two Lubbocks and an ever-changing retinue

of native peoples and scientists. On the whole, however, the story is easy to follow, and although Mithen's imaginative reconstructions will give many archaeologists pause, they will probably help non-archaeologists to visualize the richness of prehistoric life.

Covering so many sites and such a broad timeframe, *After the Ice* is always on the verge of becoming an essentially descriptive account. Mithen avoids this by focusing on the key questions that continue to provoke debate about the Pleistocene–Holocene transition. What caused the extinction of Pleistocene megafauna — climate change or human overkill? Did humans enter the New World before 20,000 BC? Why did agriculture develop in the Holocene but not earlier? In general, Mithen's answers are solidly mainstream and reflect a careful weighing of the full range of evidence and arguments that will benefit non-specialists.

What Mithen can't seem to answer, even to his own satisfaction, is whether humans are better off as the result of the cultural developments he describes and those that followed, a question he takes up in the epilogue. He would not like to live in the past as a hunter-gatherer, but questions the worth of modern conveniences bought at the price of global pollution and environmental destruction — a trade-off that he fears will worsen in the future. This part rankles, not because the issues are trivial but because archaeology provides no answers. It is enough that archaeology can tell us about the past and sharpen our understanding of what it means to be human. *After the Ice* does both. ■

Robert L. Bettinger is in the Department of Anthropology, University of California, Davis, California 95616, USA.

Thinking of apes

Intelligence of Apes and Other Rational Beings

by Duane M. Rumbaugh &

David A. Washburn

Yale University Press: 2003. 326 pp. \$35

Andrew Whiten

This intriguing book recounts a lifetime's effort by Duane Rumbaugh to establish an accurate and objective picture of the cognitive capacities of apes and other primates, spanning the past half-century's sea changes in our understanding of animal minds. When Rumbaugh began his career, the field was still dominated by a narrowly focused behaviourism that tended to reduce all learning to simple conditioning. But in the past 20 years, he has collaborated with Sue Savage-Rumbaugh to generate claims of cognitive richness in apes that are about as far from classical behaviourism as it is possible to imagine.

We are treated in this book to an honest and intimate account of Rumbaugh's struggle to square his early training and continuing commitment to behaviourism ("the only data available to our science are behaviors") and to good, objective science (through the use of automated keyboards to record data, for example) with the quirky creativity and rationality that he encountered in decades of work with chimpanzees and other primates.

Examples of this cognitive complexity include both experimental results and one-off observations of unique, innovative actions. The inherent non-replicability of the latter is of course scientifically problematic, but may nevertheless be the kind of raw material with which any scientific engagement with creative rationality must deal.

An illustration comes from an attempt to use food rewards to train the chimpanzee Lana to urinate in a pan. Rumbaugh points out that, contrary to classical conditioning theory, her response to this reinforcement did not become more vigorous and stronger; instead, she came to eke out her contributions in smaller and smaller doses, thus gaining more rewards. The key observation, however, is that when she had no urine left, she approached the pan and spat into it instead. So what Lana had learned cannot be explained as operant conditioning of her urination behaviour. Her innovation can instead be described as creative, rational and intelligent, and it is one of numerous such cases that pepper the book.

Rumbaugh's half-century scientific quest to understand the kinds of minds that lie behind such episodes may offer an education for the younger generation of researchers who can all too easily fail to appreciate the scope of the intellectual revolution that has taken place. However, Rumbaugh and



Learning curve: the chimpanzee Lana has taught scientists a great deal about intelligence in apes.

Washburn have greater ambitions: to identify a larger category of complex behaviour, which they call emergents, that goes beyond operants and respondents, integrating all three into a coherent theoretical framework, 'rational behaviourism'. No succinct definition of emergents is offered, but their scope can perhaps be glossed as behaviours for which no specific reinforcement history offers an adequate explanation. These include innovative or creative actions, such as Lana's pan-filling, along with others documented through the quantitative and statistical results of numerous experiments.

Prime examples of these are experiments on 'learning set'. In any series of such experiments, subjects are faced with the same learning contingency, such as that food is hidden under the odd one of three objects, with the objects themselves being changed for each experiment. 'Learning set' refers to the capacity to grasp the overarching rule governing the series (in this case, odd-one-out) such that, over time, successful choices are made faster with each new instantiation of the task. Rumbaugh and Washburn say that this achievement must count as an emergent. Considering this and a range of other kinds of evidence, they conclude that chimpanzees, and probably the other great apes, show the greatest manifestations of emergents, in some cases achieving a qualitative superiority over the smaller-brained monkeys tested.

The book, then, makes two key claims: that our understanding of behaviour is advanced by the concept of emergents, which is in turn nested within rational behaviourism; and that apes show both quantitative and qualitative superiority over other animals in these terms.

Should we buy these claims? On the first, I am sceptical, for several reasons. For a start, it doesn't seem clear that such phenomena as learning sets are inherently more reliant on causal reasoning or other aspects of rationality than simpler kinds of associative learning. Instead, where the latter can be described as learning about first-order patterns in the world (for example, learning

that one of three specific objects is the one to pick), the emergent can be described as learning about second-order patterns (that the game is always to pick the odd object among three, say).

My second reason is that contemporary cognitive learning theory seems to have assimilated many of the complexities that the authors worry about, such as the case of Lana learning not a specific behaviour, but about achieving a state of the world (fluid in pan). More bridges have been built between learning theory and the products of the cognitive revolution than the authors acknowledge, so it is a concern that contemporary learning theorists such as Nicholas Mackintosh and Anthony Dickinson, as well as integrators such as Sara Shettleworth, are absent from the bibliography.

I am more convinced of the relative achievements of the apes, as this is consistent with much other evidence. However, Rumbaugh and Washburn fail to cite a recent book — *Primate Cognition* by Michael Tomasello and Josep Call (Oxford University Press, 1997) — that offers a much more comprehensive overview of the kinds of experiment, such as learning set, that they highlight, but which does not find the overall evidence of qualitative ape superiority persuasive. Such alternative conclusions beg a refutation, if the case for the apes is to be made convincingly.

But perhaps the main strength of *Intelligence of Apes and Other Rational Beings* is the development of further technologies to probe cognition, most recently the use of computerized joystick-driven tasks. A string of experiments exploiting these advances has already thrown up some fascinating discoveries (ironically, perhaps, mostly in macaque monkeys rather than apes), and I predict that this may represent one of the most influential legacies of the work collated in this volume. ■

Andrew Whiten is at the School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, UK.

The cry of the food critic

Don't Worry [It's Safe to Eat]. The True Story of GM Food, BSE and Foot and Mouth

by Andrew Rowell

Earthscan: 2003. 268 pp. £16.99, \$29.95

Mark Woolhouse

Andrew Rowell believes that commercial interests are taking over science, that economic and political pressures are brought to bear on researchers, and that those with unorthodox views are subject to intimidation and vilification. He makes these serious

charges in his topical, well-publicized and widely discussed volume *Don't Worry [It's Safe to Eat]*.

The scientific and political histories of genetically modified (GM) foods provide many of the illustrations of the book's main themes, although Rowell claims that the book is not anti-GM. He also uses mad-cow disease (bovine spongiform encephalopathy, or BSE) and foot-and-mouth disease to back up his claim that critical voices are marginalized in scientific debates. He considers that, in these areas, the British public has been "betrayed" by the scientific establishment. Few escape his reprobation: the UK government, industry, the universities, the Department for Environment, Food and Rural Affairs, the Food Standards Agency, the Biotechnology and Biological Sciences Research Council (Britain's main funding agency for research in non-medical life sciences) and the Royal Society all fare badly. *Nature* is also censured, for its "fudged" retraction of a publication on GM maize.

It is easy to agree with many of the general points that Rowell makes. Scientific advice to industry or government should reflect genuine scientific uncertainties. Absence of evidence is not evidence of absence; where more research is warranted into possible public health risks or the like, it should be carried out. Research results, properly arrived at, should not be suppressed, however inconvenient they may be. Scientific controversies should be resolved by open debate, free of intimidation and harassment. If, as Rowell believes, these self-evident rules are repeatedly being broken, then that would indeed be a cause for concern.

It is the specifics that let the book down. No one could dispute that there have been heated controversies over GM food, BSE and foot-and-mouth disease, but the accounts given here are so one-sided as to put off anyone other than the most hardened conspiracy theorist. Rowell, an investigative journalist, uses a wide variety of sources — ranging from *Nature* through reports of government inquiries to the satirical magazine *Private Eye*, supplemented by numerous interviews — but he uses them selectively, sometimes perpetuating misconceptions (for example, about how models of foot-and-mouth disease were developed) in making his case. There are significant omissions too. For example, Rowell develops at length the theme that there was a collective unwillingness to determine whether BSE posed a public-health risk, without mentioning that the CJD Surveillance Unit was set up in Edinburgh in 1990 in response to the Southwood Report for just that reason.

Rowell is also unwilling to consider that unexpected (and sometimes unwelcome) findings might be criticized for no more sinister reason than that the underlying methodology is flawed. The objective of peer

Exhibition

A hobbit-forming show

This cave-troll stars in an exhibition at the Science Museum in London to celebrate the technology used to make the hugely successful films based on J. R. R. Tolkien's novel *The Lord of the Rings*.

Some film-makers would have shot the epic trilogy entirely in the computer, but director Peter Jackson used techniques ranging from advanced mathematical modelling and virtual reality to Medieval methods of armour manufacture.

Some have commented that it is hardly the business of a science museum to give space to a show about a fantasy film. In response, the museum's exhibition manager, James Rudoni, says that the exhibition is "quite simply about the science and technology behind the most incredible movie project ever undertaken. The exhibition looks behind the innovation of the film-makers."

Some 18,000 visitors went to see the exhibition in its first week alone. It runs until



11 January 2004 before moving to Singapore, Boston and Sydney.
Henry Gee
▶ http://www.sciencemuseum.org.uk/exhibitions/lordoftherings/default_flash.asp

review is not, as he seems to imply, to preserve the status quo for its own sake, but to filter out bad science. It is often true that controversial and high-profile results receive much closer scrutiny than routine results; that is hardly surprising. What matters is that such scrutiny takes place, is fair and works to generally accepted standards. In fact, much of the 'science' reported in this book appears not to have been subject to peer review at all.

Nevertheless, the book does raise the important question of how science should relate to industry and government while retaining its independence and working, as a publicly funded enterprise should, for the public good and with the public's trust. Rowell offers several suggestions in his final chapter, based on the three fine-sounding "principles" of humility, pluralism and diversity. Personally, I was too disenchanted by what had gone before to be overly impressed by these offerings; other readers may be more patient and charitable.

As this book illustrates, how science and scientists are perceived is changing in an important way. Nowadays, the pronouncements of the scientific establishment are not going to be accepted just because it is the scientific establishment. Alternative views will be sought and, if only because scientists are always ready (thankfully) to challenge main-

stream thinking, they will often be found. This is all to the good: debate is vital, and should be warmly welcomed. The point to emphasize is that it must be informed and responsible debate.

This is especially important when scientific disagreements have a direct effect on the public. For example, the recent controversy in Britain about an alleged association between the MMR vaccine (for mumps, measles and rubella) and autism is widely held to be responsible for a recent fall in vaccination coverage. This in turn has potentially serious public-health consequences of its own — a point that Rowell fails to make in a passing reference to this topic.

Ultimately, it is up to the scientists themselves to ensure that the debate is both informed and responsible. The artist James McNeill Whistler put it well in 1878: "I maintain that two and two would continue to make four, in spite of the whine of the amateur for three, or the cry of the critic for five." GM food, BSE and foot-and-mouth disease have all produced a great deal of crying and whining, but this must not goad scientists working on these subjects into making their two and two into anything other than four. ■

Mark Woolhouse is at the Centre for Infectious Diseases, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK.

Positions, please...

Niall Dillon

Rediscovered in 1900 from the research of Gregor Mendel, and named in 1909 by Wilhelm Johanssen, the gene became one of the most influential scientific concepts of the twentieth century. Yet despite its iconic power, it remains a curiously nebulous entity that defies easy definition. From the start, there was a tension between the concept of the gene as a 'unit of inheritance' — which was defined in purely operational terms as an autonomous unit that transmits specific traits through multiple generations — and the gene as a physical entity — whose position could be mapped in relation to other genes on the chromosome.

The difficulty in reconciling these two views of the gene was highlighted when Alfred H. Sturtevant and Hermann Muller separately described a new type of mutation in *Drosophila* in the 1920s and early 1930s. These mutants were the result of translocations that alter the positions of genes on the chromosome but not their physical structure. For example, placing a gene close to a compacted, transcriptionally silent region (heterochromatin) often gives rise to flies with 'variegated' phenotypes (such as patches of red and white eye colour) where the affected gene is expressed in some cells and silenced in others.

The discovery that genes could be affected by their position on the chromosome raised the question of whether the autonomous gene proposed by classical genetics actually existed. Leslie Dunn commented in 1937 that the gene was showing "signs of disappearing in a cloud of position effects"; and Richard Goldschmidt adopted the highly controversial view that the particulate gene concept should be abandoned altogether. The gene was — and still is — too useful to be discarded, but even now, 50 years after the discovery of the double-helical structure of DNA, the problem of the relationship between Johanssen's unit of inheritance and the gene as a physical structure is still not fully resolved.

A gene gives rise to a phenotype through its ability to generate an RNA or protein product. Thus the functional genetic unit must encompass not only the DNA that is transcribed into RNA, but all of the surrounding DNA sequences that regulate RNA transcription. If it is to satisfy the full definition of the particulate gene, the functional unit should also be isolated from neighbouring genes.

Transgenic assays have been used to try to determine the nature of the genetic

functional unit and to establish a physical basis for gene autonomy. Because most transgenes are subject to position effects, they can be used to search for sequences that allow them to function normally when integrated at any position on the chromosome. Candidate sequences are introduced into the transgene construct, and expression is analysed to test for sensitivity to integration position. These studies have led to two quite different perspectives on the problem.

The simplest explanation for gene autonomy would be for each gene to be flanked by barriers or insulators that isolate it from surrounding sequences. But despite intensive searches, such barriers have proved to be elusive. Where they have been found, they have turned out to be extremely diverse in structure, and often coincide with sequences that have other functions (for example, gene promoters). This heterogeneity argues against the idea that barrier effects evolved as specific and highly conserved functions. Instead, it suggests that sequences with other roles can sometimes acquire a barrier function by virtue of their location.

There is also a quite different way of looking at genetic functional units, which focuses on the interactions with transcription factors (sequence-specific DNA-binding proteins) that make a gene competent to transcribe its RNA product. If these factors are able to bind to their recognition sites and initiate chromatin remodelling despite the presence of inhibitory chromatin, then they will be able to overcome position effects. Dominance of clustered transcription-factor binding sites over negative position effects was first described by Frank Grosveld and colleagues for the human β -globin locus control region, and has since been demonstrated to operate at a number of other genes. These results portray the functional unit primarily as a dynamic information processor, comprising the transcribed region and associated factor-binding sites, which may extend over hundreds of kilobases of DNA.



Eye-opener: studies of *Drosophila* showed that a gene's position can affect its function.

Gene autonomy

Position effects continue to raise questions about the physical structure of the particulate mendelian gene.

Crucially, this type of organization allows functional units to overlap yet still remain isolated from one another if they bind to different sets of factors. As more genes are characterized in detail, it is becoming clear that overlap of genetic functional units is a widespread phenomenon.

Recent studies have also provided intriguing evidence that autonomy might be an optional feature for many genes. Microarray analysis of large numbers of genes in *Drosophila* and *Caenorhabditis elegans* showed that genes with similar patterns of expression are often grouped together on the chromosome — despite being functionally unrelated. The authors of one of these studies, Paul Spellman and Gerald Rubin, have noted that 'leaky' gene expression is often phenotypically silent. They speculate that such expression could be a frequent consequence of being located close to a highly expressed gene. If correct, this would indicate that complete isolation from neighbouring genes — instead of being an intrinsic property — evolves when it is required to prevent harmful position effects.

The particulate gene has shaped thinking in the biological sciences over the past century. But attempts to translate such a complex operational concept into a discrete physical structure with clearly defined boundaries were always likely to be problematic, and now seem doomed to failure. Instead, the gene has become a flexible entity with borders that are defined by a combination of spatial organization and location, the ability to respond specifically to a particular set of cellular signals, and the relationship between expression patterns and the final phenotypic effect.

Niall Dillon is in the Gene Regulation and Chromatin Group, MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College, Hammersmith Campus, Du Cane Road, London W12 0NN, UK.

FURTHER READING

Beurton, P., Falk, R. & Rheinberger, H.-J. (eds) *The Concept of the Gene in Development and Evolution* (Cambridge Univ. Press, Cambridge, 2000).

Dillon, N. & Sabbatini, P. *BioEssays* **22**, 657–665 (2000).

Spellman, P. T. and Rubin, G. M. *J. Biol.* **1**, 5 (2002).

Floods of record

Chris Paola

Research on two Bolivian rivers provides explanations of how and when they flood. It also gives pointers for interpreting Earth's history as recorded by the sediments left behind by flood waters.

One of the quirks of the Earth's stratigraphic recording system is that it operates in fits and starts, archiving surface conditions for a while and then suddenly switching off. The sedimentary 'record' button is evidently pretty finicky. It is sensitive to external controls, but also to the internal dynamics of the sediment-deposition system. An intriguing example of how the recording process works is given on page 493 of this issue. Aalto *et al.*¹ show how changes in the way that rainwater is delivered to a river system, driven by the El Niño/Southern Oscillation (ENSO) climate cycle in the Pacific Ocean, can switch one of the major processes of sedimentation on and off.

River flooding seems a fairly straightforward process — so much so that one could easily overlook the fact that no obvious natural law decrees how often and by what means a river should flood, or, for that matter, that it should flood at all. An alluvial river channel (one formed from its own sediment) and its adjoining flood plain are self-formed from the interplay of water, sediment and vegetation. The dynamics of flooding results from the way in which the channel–flood-plain system organizes itself. In many alluvial rivers, sedimentation associated with flooding creates 'natural levees', low ridges that run along the edges of the channel. These levees are naturally formed counterparts of the familiar artificial levees built to control flooding in populated areas. They raise the 'overtopping' threshold of the channel and in so doing create a pressure-head that helps drive water onto the flood plain once the levee is breached. So, rather than gradually inundating the flood plain, water can enter as a relatively focused, narrow flow that often carries with it a good deal of suspended sediment. Once the flow leaves the breach (the 'crevasse'), it expands and dumps its load in the form of a fan, colourfully referred to as a crevasse splay. The deposits of these fans fill much of the flood plain around the main channel with distinctive sheets, usually of sand or silt^{2–5}.

An interesting twist concerns the effects of clear water (also referred to as black water) on the flood plain. Clear water can come directly from rainfall, or it can be channel-derived water that has deposited its charge of



Figure 1 River features. The Mamore river in flood, with (inset) an example of a nascent natural levee, formed from sediment, that grows to line the river channel.

sediment elsewhere. In either case, if the flood plain is inundated with clear water, and if the channel flood wave rises slowly enough, the water pooled on the flood plain produces back-pressure that works with the levee system to contain the channel flow. This back-pressure becomes more effective as the flood plain becomes higher. In addition, inundating the flood plain with clear, particle-free water inhibits sedimentation there, instead sending the sediment on downriver⁶.

Aalto *et al.*¹ take these ideas forward. They show how clear-water effects, along with ENSO-related variability in rainfall delivery, turn flood-plain sedimentation on and off in two rivers, the Beni and the Mamore (Fig. 1). These rivers rise in the Bolivian Andes and run eastwards into the Amazon basin, depositing large amounts of sediment in a major topographical depression (the Andean foreland) as they go. Through a combination of coring, oceanographic-style data collection over a broad area of the river basin and a new ²¹⁰Pb dating technique, Aalto *et al.* show that crevasse-splay deposits in these rivers are clustered temporally, and that these clusters are associated with cold (La Niña) phases of the ENSO climate cycle.

One might think of explaining the clustering of flood-plain depositional events in terms of higher flood levels, but the authors find that flood magnitude alone

does not explain their data. They argue instead for a more subtle and interesting explanation: the controlling variable is not the size of the flood, but how quickly it rises, because of the clear-water effect outlined above. Slow rises allow time for clear water to pool on the flood plain, and the resulting back-pressure counters that of the main flood wave. With reduced pressure on the levee, there are fewer levee breaches and hence fewer splay deposits.

Aalto *et al.* stress the implications of their findings for our understanding of how climate cycles control the sequestration of sediment — along with associated nutrients and carbon — in river flood plains. But their work has a further significance. The stop-and-go nature of sedimentation leads to pervasive discontinuity in the stratigraphic record⁷. Time is missing from sedimentary sequences on all scales (imagine what it would be like if you remembered only one year of each decade, and of that year, only one of every ten days, and so on). This discontinuity gives recorded planetary (geological) time a different architecture to human time — fractal as opposed to continuous⁸.

It is easiest to measure stratigraphic discontinuity at the extremes. On short timescales (seconds to days), it can be constrained by observations of sedimentary structures, and on long timescales (typically 100,000 years and above) by rock-dating techniques. The processes that lead to discontinuity at scales between these limits are much less well understood. Aalto and colleagues' work provides one of the best-documented examples of the role that climate variability, which is thought to be strong over these intermediate timescales, could play in controlling the Earth's temperamental 'record' button.

Chris Paola is in the Department of Geology and Geophysics, and St Anthony Falls Laboratory, University of Minnesota, Minneapolis, Minnesota 55414, USA.

e-mail: cpaola@umn.edu

1. Aalto, R. *et al.* *Nature* **425**, 493–497 (2003).
2. Aslan, A. & Autin, W. J. *J. Sedim. Res.* **69**, 800–815 (1999).
3. Blum, M. D. & Törnqvist, T. *Sedimentology* **47** (suppl. 1), 2–48 (2000).
4. Slingerland, R. & Smith, N. D. *Geology* **26**, 435–438 (1998).
5. Smith, N. D., Cross, T. A., Dufficy, J. P. & Clough, S. R. *Sedimentology* **36**, 1–23 (1989).
6. Mertes, L. A. K. *Wat. Resour. Res.* **33**, 1749–1762 (1997).
7. Sadler, P. M. *J. Geol.* **89**, 569–584 (1981).
8. Plotnick, R. *J. Geol.* **94**, 885–890 (1986).

Dangerous liaisons

William R. Heath and Francis R. Carbone

Alerting the immune system to invading microorganisms is essential for effective immunity. Uric acid released by damaged cells is a danger signal that is able to notify immune cells of microbial attack.

When microorganisms invade our bodies, it is the firepower of the immune system that makes the difference between minor illness and a far worse outcome. But even the best armies are of little use without a means of detecting infiltrators. Writing on page 516 of this issue, Shi *et al.*¹ describe how they have identified something that has eluded immunologists until now — a danger signal released by damaged tissues that notifies immune cells that an invasion is under way.

Two theories have been proposed to explain how the immune system is recruited to active duty. In 1989, the late Charles Janeway proposed that immune cells might possess 'pattern-recognition receptors' — molecular microdetectors that recognize common structures on invading pathogens^{2,3}. Some years later, Ruslan Medzhitov, working in Janeway's laboratory, identified one such microdetector. This 'Toll-like receptor' was the first of ten such receptors to be discovered⁴. This discovery, together with many others, provided compelling evidence for what is sometimes called Janeway's 'stranger' hypothesis (Fig. 1a).

Unhappy with some aspects of Janeway's proposal, Polly Matzinger put forward an alternative view — the 'danger' hypothesis^{5,6}. Matzinger suggested that it is not the presence of a microbe *per se* that is important, but whether it is dangerous (because not all are). She proposed that, during microbial attack,

the resulting cellular stress or tissue damage alerts the immune system to arm its infantry (Fig. 1b). Again, immune-cell microdetectors were proposed, but Matzinger suggested that, instead of recognizing microbe parts, they recognize signs of tissue distress, or molecules that are normally found only inside cells, unless released by damage.

Although this is a compelling idea, there have been few convincing experimental data to support it. Virus-infected cells can secrete alarm signals, such as the molecule interferon- α , but whether this is a response to cellular stress — a danger signal — or results from the detection of characteristic viral structures such as double-stranded RNA — a known 'stranger' signal⁴ — is unclear. Similarly, proteins called heat-shock proteins have been reported to act as cellular alarm signals when released from damaged cells⁷, but whether they raise the alarm because they are true danger signals or because these sticky proteins easily associate with contaminating microbial products (stranger signals) is contentious.

Despite this lack of convincing evidence for a bona fide danger signal, several groups have reported that undefined cellular material^{8,9}, including cytoplasm¹⁰ (the intracellular molecular soup), can activate the immune system. Now Shi *et al.*¹ have identified a specific molecule in cytoplasm that acts as an alarm. The authors fractionated the components of cytoplasm using chromatography. Then, to find out which of the fractions

contained the alarm molecule, they injected material from each fraction into mice, together with latex beads covered with the HIV protein gp120. One of the fractions contained most of the alarm signal — as measured by its ability to boost the immune response to gp120 — and the major component of this fraction was uric acid. This is a degradation product of nucleotides, which the cell uses in various ways, for example to make genetic material (DNA and RNA) and in energy transfer (ATP).

To confirm that uric acid was indeed the cytoplasmic danger signal, the authors showed that a highly purified form of the chemical could also act as an alarm. And uricase, an enzyme that degrades uric acid, reduced the ability of cytoplasm to boost the immune response. Shi *et al.* also showed that both cytoplasm and purified uric acid activated a central immune player, the dendritic cell.

Finally, the authors found that uric acid activated dendritic cells only when in its crystal form (Fig. 1b). This might at first seem surprising, but it perhaps fits with the fact that soluble uric acid is a normal by-product of nucleic-acid degradation and is usually present at high concentrations in blood and in the fluid between tissues. It is therefore important that immune cells are oblivious to soluble uric acid, or else a constant alarm would ring. Because the normal concentration of uric acid is near-saturating, crystals might easily form in tissues when a small excess of this chemical is released during cellular damage. So using uric acid as a danger signal might allow the immune system to be extremely sensitive to local cell damage. Interestingly, the formation of uric-acid crystals in joints causes gout, an acute form of arthritis associated with painful inflammation. Perhaps the finding that these crystals stimulate dendritic cells helps to explain the inflammatory effects of uric acid.

Shi and colleagues' findings¹ provide compelling support for Matzinger's danger theory^{5,6}. So far, however, we only know that uric acid can boost immunity to particulate antigens — in this case, latex beads coated with ovalbumin or gp120. It will be important to see whether uric acid functions more broadly in developing immunity to viral and bacterial infections, and whether such danger signals operate in other types of immune response that are not known to involve overt microbial infections, such as certain autoimmune responses, antitumour responses or transplant rejection.

Where does the new study leave Janeway's stranger hypothesis? It seems likely that there is more than one pathway to effective immunity, and that the mechanism for detecting cellular damage described by Shi *et al.* complements, rather than replaces, Janeway's model. The immune system probably

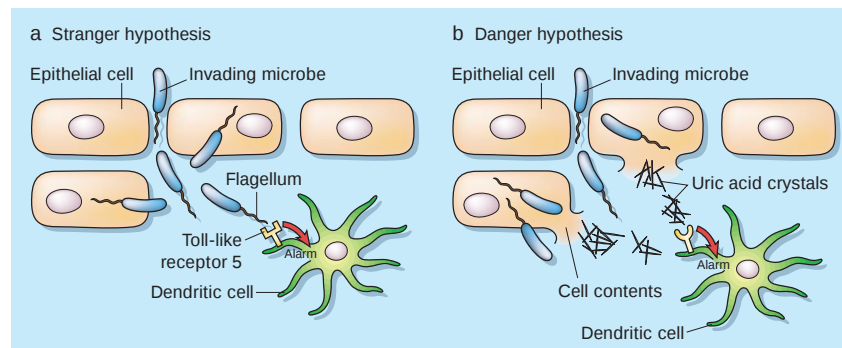


Figure 1 Strangers and danger. Two different theories describe how the immune system is alerted to respond to microbial attack. a. In the 'stranger' model^{2,3}, microdetectors on the surface of immune cells recognize common molecular patterns on the surface of microbes. For example, Toll-like receptor 5 on the surface of dendritic cells signals an alarm when it recognizes a protein component of the flagella of some bacteria⁴. b. Alternatively, the 'danger' hypothesis^{5,6} proposes that the immune system is alerted to the tissue damage associated with microbial infection. Shi *et al.*¹ provide evidence for this second view, showing that damaged cells release uric acid that forms crystals capable of activating dendritic cells. The now-compelling evidence for both hypotheses suggests that the immune system scans for signs of both danger and strangers.

ensures vigilance by scanning for both danger and strangers. ■

William R. Heath is in the Division of Immunology, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia.
e-mail: heath@wehi.edu.au

Francis R. Carbone is in the Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria 3010, Australia.
e-mail: fcarbone@unimelb.edu.au

1. Shi, Y., Evans, J. E. & Rock, K. L. *Nature* **425**, 516–521 (2003).

2. Janeway, C. A. Jr *Cold Spring Harb. Symp. Quant. Biol.* **54**, 1–13 (1989).
3. Medzhitov, R. & Janeway, C. A. Jr *Science* **296**, 298–300 (2002).
4. Takeda, K., Kaisho, T. & Akira, S. *Annu. Rev. Immunol.* **21**, 335–376 (2003).
5. Matzinger, P. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
6. Matzinger, P. *Science* **296**, 301–305 (2002).
7. Srivastava, P. *Nature Rev. Immunol.* **2**, 185–194 (2002).
8. Sauter, B. et al. *J. Exp. Med.* **191**, 423–434 (2000).
9. Gallucci, S., Lolkema, M. & Matzinger, P. *Nature Med.* **5**, 1249–1255 (1999).
10. Shi, Y., Zheng, W. & Rock, K. L. *Proc. Natl Acad. Sci. USA* **97**, 14590–14595 (2000).

Developmental biology

Twisting the body into shape

Ian C. Scott and Didier Y. R. Stainier

Molecular signals are not the only forces that pattern and shape the developing embryo. Mechanical stresses sensed by cells also seem to be involved in creating the body plan.

Embryogenesis — the process by which a single cell gives rise to a multicellular organism — is one of nature's wonders. As cells increase in number, they engage in a series of complex movements that substantially alter their relative positions in the embryo. Patterning information that specifies anterior–posterior (head to tail) and dorsal–ventral (back to front) orientation is superimposed on this cellular framework, leading to the development of the early body plan. Many of the molecular signals that convey this patterning information

have been identified. Now, writing in *Current Biology*, Emmanuel Farge¹ provides evidence that the mechanical forces generated by cellular migration can also play a role in shaping the embryo.

During the early stages of embryogenesis, a ball of cells is transformed into a multi-layered structure, as cells invaginate from the surface and move inwards. This process generates the three 'germ layers', which go on to form all the organs of the body. Subsequent migrations, and changes in the cells' shape and adhesive properties, further alter the

embryo's form. As the cells are linked to an extracellular matrix, they are subjected to forces exerted both by their own movements and by those occurring in other regions of the embryo. Mechanical stresses have been shown to affect gene expression in several cell types, most notably those lining the blood vessels². So could the forces resulting from widespread cellular movements during embryogenesis affect the expression of the developmental genes required to set up the embryo's body plan?

To explore this possibility, Farge¹ mechanically compressed embryos of the fruitfly *Drosophila*, and then examined the expression of specific developmental genes. A gene called *twist* — involved in dorsal–ventral patterning³ — is normally expressed only in the ventral region of the embryo, but Farge found that, within eight minutes of compression, *twist* expression had spread to encompass the entire embryo. Intriguingly, embryos in which the dorsal–ventral patterning system had been eliminated genetically also showed this expanded pattern of *twist* expression when compressed. So these results suggested that a pathway that is responsive to mechanical stress could affect patterning in the embryo.

But how is mechanical force translated into a change in gene expression? One protein that might couple these events is β -catenin, which is involved in both cell adhesion and gene activation⁴. Indeed, Farge found that compressing the embryo induced β -catenin to move into the cell nucleus — the site of gene activation. And inhibiting β -catenin, either by mutation or by overproducing proteins that interfere with β -catenin's gene-activating function, suppressed the stress-induced expression of *twist*.

So it seems that, through β -catenin, mechanical forces applied to the developing embryo can indeed affect the expression of developmental genes. But these experiments involved an artificial, external compression that does not occur during embryogenesis. To determine whether mechanical stresses inside the embryo contribute to its patterning, Farge examined a group of cells that are fated to form the most anterior portion of the digestive tract, called the stomodeum. He chose these cells because they lie between tissues that undergo extensive movements during embryogenesis: the posterior mesoderm, which lengthens and pushes on the anterior part of the embryo, and the invaginating anterior mesoderm and foregut. Farge found that as these tissues moved, the shape of the stomodeal precursor cells seemed to become compressed, and this compression coincided with increasing *twist* expression. Furthermore, as seen in the mechanically compressed embryos, the increase in *twist* expression required β -catenin activity in the nucleus.

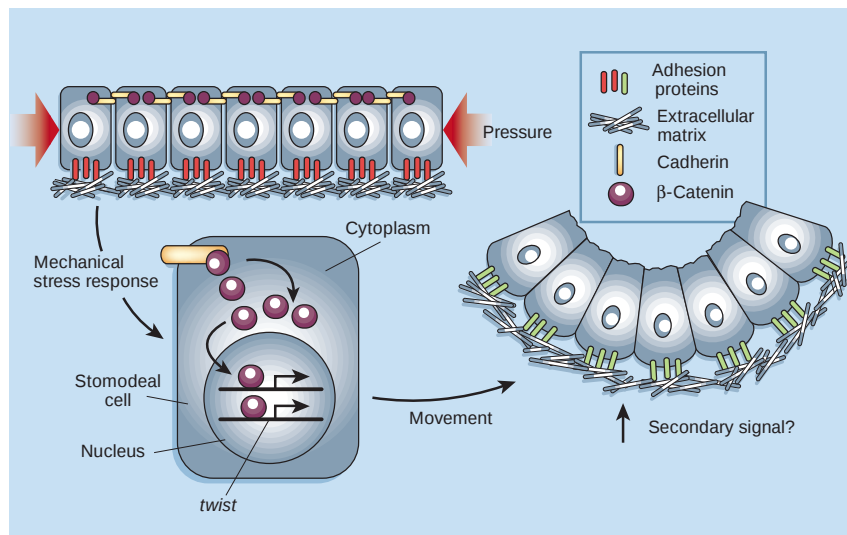


Figure 1 The mechanics of shaping the digestive tract. New work by Farge¹ shows that mechanical compression of stomodeal precursor cells (which will form the part of the digestive tract called the stomodeum) can affect the expression of the *twist* gene. Pressure causes the β -catenin protein to move from the cell membrane (where it associates with cadherin) to the cytoplasm, increasing its concentration there. This movement in turn allows β -catenin to accumulate in the nucleus, where it activates *twist* and probably other developmental genes. *twist* is required for the invagination and development of the stomodeum. Possible 'downstream' responses may include changes in the contacts made between cell adhesion molecules and the extracellular matrix, facilitating invagination. The reaction to mechanical stress might also be 'permissive', allowing the stomodeal cells to interpret a secondary signal originating from surrounding tissues.

Was the increase in *twist* expression in the stomodeal precursor cells really caused by movements of the surrounding tissues, or was it just a coincidence? To answer this question, Farge disrupted the forces exerted by the lengthening posterior mesoderm, either by inhibiting the lengthening itself or by eliminating the cells that link this tissue to the stomodeal precursor cells. In both cases, the stomodeal precursors no longer seemed to be compressed and showed no increase in *twist* expression. But, dramatically, when Farge compressed these cells using a metal probe, *twist* expression was activated. As this gene is known to be required for the stomodeum to invaginate and assume its normal shape⁵, it now seems that mechanical stress, through the activation of *twist*, might directly regulate the development of this tissue (Fig. 1).

The effect of developmental genes on tissue movements during embryogenesis is being studied intensely. Farge's work shows that the mechanical forces generated by these movements can in turn regulate the expression of developmental genes. These forces have not been studied in detail and closer scrutiny is warranted⁶. For example, it will be interesting to investigate whether the stress-response pathway exists in other developmental contexts, such as the looping of the heart and gut tubes, where mechanical forces are probably involved. Another question is whether cells sense mechanical forces differently according to the nature of their contacts with the extracellular matrix. Some adult tissues may also experience mechanical stress — such as pressure from the growth of a tumour — so it will also be important to investigate how these stresses are involved in disease.

Current interpretations of how some genes function may also have to be revised in the light of whether or not they influence the generation of mechanical force, or the response to it. For example, it was previously known that β -catenin moves to the nucleus as a result of intracellular signalling triggered by the Wnt protein, but Farge has now shown that it can also relocate in response to mechanical stress. In the absence of either stimulus, β -catenin is found in two locations in the cell — near the cell surface, where mechanical forces could be sensed, and in the cytoplasm. So β -catenin function might have been co-opted to respond to two different stimuli.

The new study¹ prompts many other questions. How are mechanical signals transduced in other tissues? Is β -catenin a key player in all cell types? What targets other than *twist* are regulated? As Farge showed, artificial pressure can trigger β -catenin to move to the nucleus of stomodeal precursors and activate *twist*. Is this response sufficient to induce the invaginations required for the stomodeum to develop? Alternatively, is the

response to pressure simply permissive, preparing the cells to respond to other inputs that induce movement? And what determines how a cell will move in response to these inputs? One thing is certain — examining the interplay between mechanical and molecular signals should open up new avenues in the study of many biological processes, including embryogenesis. ■

Ian C. Scott and Didier Y. R. Stainier are in the Department of Biochemistry and Biophysics, University of California, San Francisco,

California 94143-0448, USA.

e-mails: ianjr88@itsa.ucsf.edu

didier_stainier@biochem.ucsf.edu

1. Farge, E. *Curr. Biol.* **13**, 1365–1377 (2003).
2. Galbraith, C. G. & Sheetz, M. P. *Curr. Opin. Cell Biol.* **10**, 566–571 (1998).
3. Ray, R. P., Arora, K., Nüsslein-Volhard, C. & Gelbart, W. M. *Development* **113**, 35–54 (1991).
4. Sanson, B., White, P. & Vincent, J. P. *Nature* **383**, 627–630 (1996).
5. Reuter, R., Grunewald, B. & Leptin, M. *Development* **119**, 1135–1145 (1993).
6. Keller, R., Davidson, L. A. & Shook, D. R. *Differentiation* **71**, 171–205 (2003).

Chemistry

Mirrors in Flatland

Rasmita Raval

Many molecules exist in two mirror-image forms, which have different biological properties. A new way of creating solid chiral surfaces might make it easier to synthesize and purify only one of the mirror forms.

Chirality is widespread in nature and is fundamental to life. It arises from straightforward geometry — any object that lacks inverse symmetry can exist in two distinguishable mirror images, called enantiomers. For example, our hands are chiral: the right hand is an enantiomer of the left hand, and neither can be superimposed on the other by translation or rotation. At a molecular level, right-handed and left-handed compounds often have very different effects, so it is crucial for industry to produce pure enantiomeric forms. On page 490 of this issue, Switzer *et al.*¹ describe a way of creating a catalytically active chiral solid that could make the production of pure mirror images much easier.

Louis Pasteur introduced the concept of molecular chirality in 1848, when he observed that crystals of the chemical sodium ammonium tartrate tetrahydrate can form left-handed and right-handed structures². Since then, chirality has been the cornerstone of several scientific advances, from the deduction that carbon atoms possess a tetrahedral arrangement of bonds^{3,4}, to the realization that terrestrial life-forms have evolved to make use of right-handed sugars and left-handed amino acids.

The inherent chirality of living systems dictates extraordinary specificity in the recognition of chiral molecules, so that a molecule and its mirror image, whether it is a pharmaceutical, an insecticide, a herbicide, a flavour or a fragrance, will almost always elicit different biological effects. This specificity presents a problem for the industrial synthesis of these compounds — chemists must control the three-dimensional spatial arrangements adopted by their products so that only the required enantiomer is produced. Underpinning this multibillion-pound global industry are chiral catalysts⁵

that promote 'enantiospecific' reactions in which only one of the mirror images is formed. Most of these reactions are homogeneous — the reactants and the catalyst exist in the same phase (generally in solution). Heterogeneous catalysis, where the catalyst is in a different phase (usually solid) from the reactants, is a fledgling technology but is often a more active and robust system that enables the products of the reaction to be separated from the catalysts much more easily. Central to this new process is the ability to incorporate chirality in catalytic solids and surfaces.

So how are chiral surfaces formed on symmetric, achiral solids? Although prevalent in the living, organic world, chirality is rarely found in catalytically active inorganic materials and surfaces. But this property can be transmitted from organic molecules to inorganic systems in a number of ways^{6–8}. For example, if organic 'right-handed' (*R,R*)-tartaric acid enantiomers are adsorbed onto the symmetric, achiral surface of solid copper, the tartaric acid molecules assemble into a chiral template on the copper surface. This exposes chiral channels within which the copper atoms are available to react with other molecules⁶. The entire assembly can assume the mirror image by adsorbing the 'left-handed' (*S,S*)-tartaric acid enantiomer instead (Fig. 1, overleaf).

As well as simply providing a template on a solid surface, adsorbed molecules can also induce the large-scale reorganization of metal surfaces to expose chiral facets. This, for example, is how the organic amino acid *S*-lysine transmits chirality to the surface atoms of solid copper⁸. And surface adsorption can also affect three-dimensional chirality^{9,10} — symmetric calcite crystals can adopt asymmetric, chiral shapes when grown in the presence of enantiomers of the amino acid aspartic acid¹⁰.

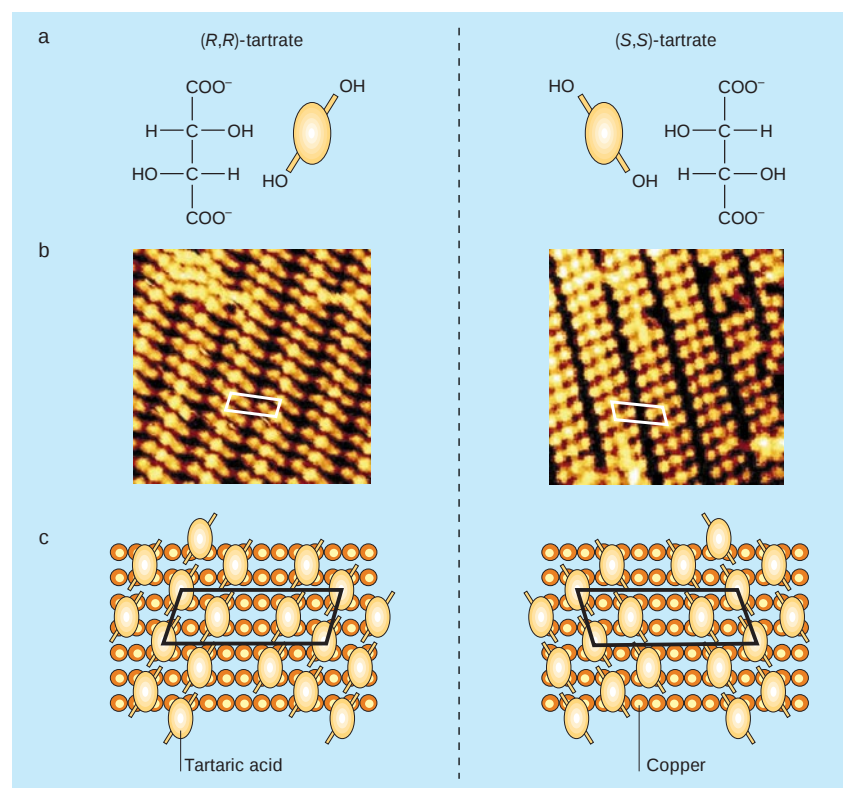


Figure 1 Bestowing chirality. A single monolayer of organic, chiral molecules adsorbed on an inorganic, achiral solid support can bestow two-dimensional chirality on the underlying inorganic surface. **a**, For example, tartaric acid can exist in two mirror-image forms, or enantiomers: (R,R)-tartaric acid and (S,S)-tartaric acid. **b**, Adsorption of pure (R,R)-tartaric acid on the surface of achiral copper creates a chiral template that destroys the mirror symmetry planes of the copper surface and exposes the copper atoms in chiral spaces. **c**, The chiral surface created by (R,R)-tartaric acid is the mirror image of that created by (S,S)-tartaric acid. The region boxed in black corresponds to the white box, with dimensions $23 \text{ \AA} \times 7.7 \text{ \AA}$, in the scanning tunnelling micrograph in **b**. Switzer *et al.*¹ have now found an alternative method of inducing chirality in an inorganic solid.

The goal of many of these strategies to produce chiral solids is to create a catalytically active chiral surface. Switzer *et al.*¹ have now achieved this goal, using an alternative method of inducing chirality in a normally achiral, inorganic material. These authors show that electrochemical methods can be used to deposit a catalytically active chiral film on an achiral solid support.

Switzer *et al.* placed a symmetric gold crystal in a standard electrochemical cell containing a solution of copper ions, tartrate ions and sodium hydroxide. By holding the gold crystal at a certain potential, a copper-oxide film was deposited on its surface. X-ray diffraction methods revealed that the deposited film was chiral — it had no mirror or inversion elements. And the chirality of the film was determined by the nature of the tartrate enantiomers present in the solution: (R,R)-tartrate yielded one structure while (S,S)-tartrate produced the mirror image. Importantly, Switzer *et al.* showed that the chirality of the film was able to preferentially catalyse the oxidation of specific tartrate enantiomers — the copper-oxide surface grown in the presence of (R,R)-tartrate was more effective

at oxidizing (R,R)-tartrate ions, whereas the surface deposited in the presence of (S,S)-tartrate was more effective at oxidizing (S,S)-tartrate ions.

The process described by Switzer *et al.*¹ is an appealing alternative way of creating

materials that combine catalytic activity and chiral selectivity. Their new technology is likely to advance the search for chirally sensitive and discriminating materials that could be used for electrochemical sensors, chromatographic separation, enantiospecific catalysis and nonlinear optics.

The challenge now for fundamental science is to understand, at the molecular and atomic levels, how the presence of organic chiral molecules disrupts the symmetry of growing inorganic crystals. Switzer *et al.* report that the chirality of their two-dimensional film of copper oxide is determined at the initial stages of deposition. And Orme *et al.*¹⁰ have suggested that the chiral growth of three-dimensional calcite crystals is determined by the adsorption of aspartic acid to the calcite surface. So the nanoscale imprinting of chirality in two dimensions could hold the key to creating macroscopic three-dimensional chiral structures. Sophisticated surface spectroscopies and imaging techniques have enabled certain manifestations of two-dimensional chirality to be mapped (Fig. 1b). But although chirality has a history of more than 150 years, we are only just taking the first steps towards understanding chirality at surfaces.

Rasmata Raval is at the Surface Science Research Centre and the Department of Chemistry, University of Liverpool, Liverpool L69 3BX, UK.

e-mail: r.raval@liverpool.ac.uk

- Switzer, J. A., Kothari, H. M., Poizat, P., Nakanishi, S. & Bohannan, E. W. *Nature* **425**, 490–493 (2003).
- Pasteur, L. *Ann. Phys.* **24**, 442 (1848).
- LeBel, J. A. *Bull. Soc. Chim. Fr.* **22**, 337 (1874).
- Van't Hoff, J. H. *Arch. Neerl. Sci. Exactes Nat.* **9**, 445 (1874).
- Sheldon, R. A. *Chirotechnology* (Dekker, New York, 1993).
- Ortega Lorenzo, M., Baddeley, C. J., Muryn, C. & Raval, R. *Nature* **404**, 376–379 (2000).
- Kühnle, A., Linderer, T. R., Hammer, B. & Besenbacher, F. *Nature* **415**, 891–893 (2002).
- Zhao, X. J. *Am. Chem. Soc.* **122**, 12584–12585 (2000).
- Addadi, L., Berkovitch-Yellin, Z., Weissbuch, I., Lahav, M. & Leiserowitz, L. *Topics Stereochem.* **16**, 1–85 (1986).
- Orme, C. A. *et al. Nature* **411**, 775–779 (2001).

Neurobiology

Backchat at the synapse

Patricia C. Salinas

At synapses, nerve cells release neurotransmitters, which affect other nerve cells or muscles. Studies of how muscles in turn influence neurotransmitter release hint at how synapses adapt to changes in use.

Synapses — points of contact between a nerve cell and its target — are dynamic structures that can grow, shrink and change their properties according to their firing history. Crosstalk between the 'presynaptic' neuron (which releases neurotransmitter molecules when stimulated) and its 'postsynaptic' target (which responds to the neurotransmitters) is crucial

for the formation, maturation and refinement of synapses. Discovering the signals — besides neurotransmitters — that are used for this crosstalk at developing synapses will shed light on processes such as learning and memory. Hence the importance of two papers in *Neuron*, in which McCabe *et al.*¹ and Haghighi *et al.*² describe their studies of a new system that is involved in synaptic

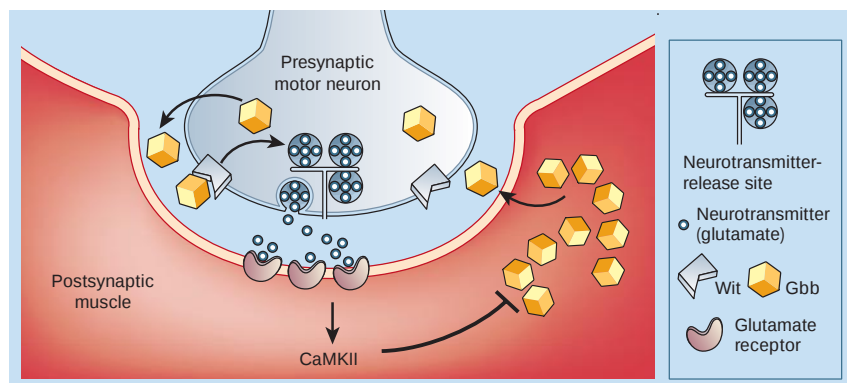


Figure 1 Retrograde signalling at the fruitfly neuromuscular junction. The findings of McCabe *et al.*¹ and Haghighi *et al.*² suggest the following model. Release of the neurotransmitter glutamate by the presynaptic neuron results in the activation of glutamate receptors in the postsynaptic muscle cell. This in turn activates the CaMKII enzyme, which blocks the expression, secretion or function of a molecule of the bone morphogenetic protein (BMP) family, possibly Gbb. When this BMP signal is released, it acts retrogradely on the presynaptic neuron via the receptor protein Wit to increase neurotransmitter release. When the signal is blocked, neurotransmitter release is reduced. Some Gbb is also released from the presynaptic neuron. The existence of this negative-feedback loop could explain how neurotransmitter release can be maintained at a steady level.

conversations. Here, a well-known class of signalling proteins carries information backwards — from the postsynaptic to the presynaptic cell.

The papers^{1,2}, which have several authors in common, focus on the neuromuscular junction (NMJ) — a connection between motor neurons and muscle — in the fruitfly *Drosophila melanogaster*. The NMJ uses glutamate as a neurotransmitter, and has similar characteristics to excitatory synapses in the vertebrate brain. It is known that backwards (retrograde) signals regulate synaptic growth and homeostasis (maintenance of a steady state) at the *Drosophila* NMJ^{3,4}. And last year, the *wishful thinking* (*wit*) gene, which encodes a receptor for signalling proteins of the bone morphogenetic protein (BMP) class, was found to regulate both synaptic growth and neurotransmitter release at the presynaptic cell^{5,6}. This finding led to the hypothesis that a member of the BMP family functions as a retrograde signal at the NMJ.

Now McCabe and co-workers¹ have found that the *glass bottom boat* (*gbb*) gene — a relative of a vertebrate BMP — is required for proper synaptic growth and transmission. Moreover, the defects seen in flies with mutations in *gbb* are similar to those observed in *wit* mutants⁵, including abnormalities in neurotransmitter release. McCabe *et al.* also show that signalling by the Gbb protein in a cell-culture system involves Wit, hinting that Wit is a Gbb receptor at the NMJ.

The authors go on to show that *gbb* messenger RNA (an indicator of gene expression) accumulates in muscles during the formation of the NMJ, but is also produced in motor neurons. So they tested the relative importance of Gbb from muscles

and from motor neurons. They found that Gbb expression in either tissue could partially overcome the defect in neurotransmitter release seen in *gbb* mutants. But the protein needed to be expressed in both tissues to completely restore synaptic function. Encouragingly, muscle-derived Gbb also restored the production of indicators of BMP signalling in the motor neuron, whereas neuron-derived Gbb was less effective. As the Wit protein is both expressed in the presynaptic motor neuron and required for presynaptic function^{5,6}, these findings imply that Gbb is likely to function retrogradely — acting, via Wit, on the presynaptic nerve cell. One explanation for McCabe and colleagues' observations is that the overall concentration of Gbb in the synaptic cleft — the space between the pre- and postsynaptic cells — regulates the presynaptic cell: some Gbb may be presynaptically derived, but most may come from muscle.

Haghighi *et al.*² provide further evidence that a retrograde signal controls NMJ transmission, and examine how the signal is regulated. Previous studies found that neuronal firing at the NMJ, and consequent activity of glutamate receptors in the postsynaptic muscle, regulates a retrograde signal⁷. Haghighi *et al.* reasoned that, because activation by glutamate triggers the entry of calcium ions into cells, Ca²⁺-sensitive enzymes such as calcium/calmodulin-dependent protein kinase II (CaMKII) could be involved in controlling the retrograde signal. CaMKII was known to be required for synaptic 'memory'⁸, but its role in synaptic growth and modulation during development had not been documented. Haghighi *et al.* found that inhibiting CaMKII in muscles increased presynaptic neurotransmitter release and the number of neurotransmitter-

release sites in the presynaptic cell. Conversely, activating CaMKII decreased both of these parameters. So CaMKII activity in the postsynaptic muscle cell regulates a muscle-derived signal that functions retrogradely to control presynaptic structure and function (Fig. 1).

The mounting evidence that this retrograde signal is a BMP-like molecule led Haghighi *et al.* to test whether the effects of CaMKII on neurotransmitter release require BMP signalling. They found that inhibiting CaMKII failed to stimulate neurotransmitter release by a presynaptic cell if the Wit protein was missing. This exciting finding hints that CaMKII regulates the expression, secretion or function of a BMP-family protein from muscle. Is this protein Gbb? It is a strong candidate, but the effect of CaMKII activity on Gbb remains to be tested.

One question that now begs an answer concerns the fact that although changes in postsynaptic CaMKII activity regulate presynaptic neurotransmitter release, CaMKII does not seem to affect synaptic growth. Yet Gbb and Wit are required for both neurotransmitter release and synaptic growth. Why this paradox? Perhaps CaMKII inhibition also regulates molecules that inhibit the growth-promoting activity of the retrograde signal. Further studies are needed to resolve this issue.

The newly discovered postsynaptic role for CaMKII — in regulating a retrograde signal that in turn controls presynaptic function — has profound implications for synaptic homeostasis: neurons can respond to changes in their activity. For example, inactivity of a synapse over a period of time can increase its strength and size. Conversely, when neurons are excessively stimulated, synaptic strength decreases so that neuronal excitability remains at normal levels^{9,10}. The mechanisms that control this compensation process have previously been poorly understood, but the new studies^{1,2} provide a possible solution. Thus, increased activity of the presynaptic cell at the NMJ results in increased levels of glutamate release; when this neurotransmitter binds to the postsynaptic cell, it results in a greater influx of Ca²⁺ ions. This, in turn, increases CaMKII activity, which somehow decreases the activity or level of the retrograde signal, resulting in less neurotransmitter release. In other words, increased neurotransmitter release results in a negative-feedback loop that ensures a specific level of synaptic transmission.

Because of the similarities between the *Drosophila* NMJ and excitatory glutamate-dependent synapses in vertebrate brains, these studies may have uncovered a general strategy by which synaptic structure and strength can be maintained. The new findings are probably just one piece of the puzzle. But CaMKII and Gbb are now

in the spotlight: can our excited synapses comprehend them?

Patricia C. Salinas is in the Department of Biological Sciences, Imperial College London, London SW8 2AZ, UK.

e-mail: p.salinas@imperial.ac.uk

- McCabe, B. D. et al. *Neuron* **39**, 241–254 (2003).
- Haghighi, A. P. et al. *Neuron* **39**, 255–267 (2003).
- DiAntonio, A., Petersen, S. A., Heckmann, M. & Goodman, C. S. *J. Neurosci.* **19**, 3023–3032 (1999).

- Kazama, H., Morimoto-Tanifuji, T. & Nose, A. *Neuroscience* **117**, 615–625 (2003).
- Aberle, H. et al. *Neuron* **33**, 545–558 (2002).
- Marques, G. et al. *Neuron* **33**, 529–543 (2002).
- Petersen, S. A., Fetter, R. D., Noordermeer, J. N., Goodman, C. S. & DiAntonio, A. *Neuron* **19**, 1237–1248 (1997).
- Lisman, J., Schulman, H. & Cline, H. *Nature Rev. Neurosci.* **3**, 175–190 (2002).
- Turrigiano, G. G. & Nelson, S. B. *Curr. Opin. Neurobiol.* **10**, 358–364 (2000).
- Davis, G. W. & Bezprozvany, I. *Annu. Rev. Physiol.* **63**, 847–869 (2001).

Materials science

Molecules squeezed and stroked

Steve Granick, Zhiquan Lin & Sung Chul Bae

Soft matter is often found in tight spots. A study shows that tangled chain-like molecules, squeezed between solid surfaces and stroked by sliding, might become exceptionally ordered.

How does a pen write on paper? As you blink your eyes, how do tears lubricate your cornea? As you type, how does the hard disk store information? All of these apparently disparate actions share a central feature — surfaces, separated by fluids, are squeezed together and the molecules sandwiched between are stroked by sliding. Sliding is one of the most common processes of the everyday world, but we know little about the interplay between surfaces and their confined lubricants.

Shinji Yamada has now looked at the effect of sliding on the behaviour of the silicone oil polymer PDMS — a flexible threadlike chain molecule that normally exists in a tangled, randomly coiled form. Writing in *Langmuir*¹, he reports that the degree of friction between two sliding surfaces depends on the number of layers of PDMS sandwiched between. Remarkably, each layer is the width of the polymer backbone, indicating that sliding might have ordered the polymer chains.

To look at the behaviour of confined PDMS during sliding, Yamada sandwiched a film of the polymer between two layers of solid mica. He then measured the kinetic friction produced as the two surfaces were squeezed and simultaneously slid over one another. Surprisingly, he was able to compress the confined molecular film during squeezing and sliding to thicknesses equalling only two, three or four molecular layers. And the number of layers affected the total friction between the sliding surfaces. When the polymer was squeezed to a thickness of three or four molecular layers, the middle layers also seemed to slide, resulting in low friction. But when the thickness of the film was squeezed to the width of only two molecular layers, the degree of friction increased abruptly by six to eight times (Fig. 1).

The finding that PDMS can form such thin layers in films is surprising, because

decades of research have led to the view that long-chain polymers form much thicker conformations — instead of aligning in flat layers they coil randomly, rather like tangled spaghetti². The polymer that Yamada looked at is about 1,000 repeat-units long, so, if it forms a random coil, the polymer film should be about 10 nm thick. But Yamada describes PDMS films that were an order of magnitude thinner than this.

A speculative but plausible explanation for this discrepancy is that sliding might have oriented the polymer chains into discrete molecular layers, each layer having the thickness of the polymer backbone (Fig. 1). Extreme deformations are most likely to happen in thin films of soft materials because the movements of these molecules are retarded by confinement and so they are deformed more rapidly than they can re-equilibrate³. The difference between polymers at rest, or at equilibrium, and polymers

undergoing deformation during sliding, is considerable.

Why are studies such as Yamada's important? If the model shown in Fig. 1 is correct, measuring the increases in friction under different degrees of compression could provide clues about how precisely a polymer is oriented by sliding. Studies of confined molecules might also have more practical implications. The ability to manipulate giant molecules under confinement may enable new kinds of nanoscale structures to be fabricated. For example, the use of conducting polymers in display devices has been limited by an inability to produce sufficient order along the polymer backbones⁴. If sliding can orient other large polymers, besides PDMS, the technique described by Yamada might be a useful route towards enhancing order in thin films of these technologically useful molecules.

But how might sliding orient giant polymers? Computer simulations and theory provide clues⁵ but no definite answers. Imaging the local density of molecules in ultrathin films and the local orientation of individual chain segments are technical hurdles that must be overcome before these questions can be answered. The most advanced synchrotron X-ray-reflectivity measurements have begun to yield images of confined molecules at rest⁶ but not yet during sliding. The interaction of visible light with matter⁷ — for example, Raman and infrared imaging — has impressive potential, but this technology is at an early stage. The problem with using these techniques to image confined molecules is distinguishing the signal (the fluid monolayer) from all the noise (the sliding surfaces and the solids beneath them).

Another task will be to distinguish the behaviour of individual confined molecules from the average behaviour. Biophysics has been invigorated by single-molecule

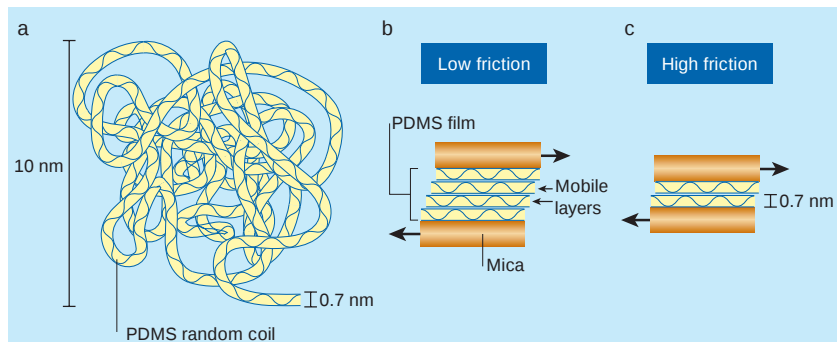


Figure 1 Squeezing out the tangles. a, The polymer PDMS usually exists in a random coil conformation. b, c, Yamada¹ has shown that PDMS might be ordered into discrete molecular layers — each layer the width of the polymer backbone — by sliding when it is confined between two atomically smooth mica surfaces. Compressing the sandwiched film to a thickness of four backbone widths produces low friction between the sliding surfaces, as the middle layers are mobile. When the film is compressed to a thickness of two backbone widths, the magnitude of friction abruptly increases. A challenge for materials scientists is to augment the macroscopic friction measurements with local measurements of molecule density, alignment, mobility and energy flow between chemical bonds as the molecules slide.

approaches to studying processes such as enzyme turnover. Applying a single-molecule approach to thin films of soft matter, one study showed that the average friction masked an unexpected variety of molecular movement⁸.

Yamada¹ has provided a conceptual model of how sliding might order large polymer backbones. The challenge now is to put flesh on this model by revealing the mechanisms underlying this phenomenon. The word friction evokes the triviality of changing the oil in one's automobile, yet, as David Tabor emphasized ten years ago⁹, a fundamental issue must be addressed before we can fully understand friction and sliding — how is energy dissipated when surfaces slide over one another? What is the roadmap of energy flow as molecules slide past one another? When these questions are answered, the word 'friction' will have real substance, and models of friction will be

predictive. Then we will have to apply our understanding of friction to issues such as biolubrication¹⁰. Machinery wears out, as we all know. So do eyes and knees. ■

Steve Granick is in the Departments of Materials Science and Engineering, of Chemistry, and of Physics, and Zhiqun Lin and Sung Chul Bae are in the Materials Research Laboratory, University of Illinois at Urbana-Champaign, Illinois 61821, USA. e-mail: sgranick@uiuc.edu

1. Yamada, S. *Langmuir* **19**, 7399–7405 (2003).
2. Jones, R. A. L. & Richards, R. W. *Polymers at Surfaces and Interfaces* (Cambridge Univ. Press, 1999).
3. Granick, S. *Phys. Today* **52**, 26–31 (1999).
4. Scherf, U., Riechel, S., Lemmer, U. & Mahr, R. F. *Curr. Opin. Solid State Matter Sci.* **5**, 143–154 (2003).
5. Bhushan, B., Israelachvili, J. N. & Landman, U. *Nature* **374**, 607–616 (1995).
6. Seeck, O. H. et al. *Europhys. Lett.* **60**, 376–382 (2002).
7. Salmeron, M. *Tribol. Lett.* **10**, 69–79 (2001).
8. Mukhopadhyay, A., Zhao, J., Bae, S.-C. & Granick, S. *Phys. Rev. Lett.* **89**, 136102 (2002).
9. Tabor, D., Singer, I. L. & Pollack, H. M. (eds) *NATO Advanced Study Institute on Fundamentals of Friction* (Kluwer, Dordrecht, 1992).
10. Zhu, Y. & Granick, S. *Macromolecules* **36**, 973–976 (2003).

Developmental biology

Partners united

Matthew Freeman

There is often more than one way of cracking a scientific problem. Two views of one question have led to the marriage of two signalling proteins in search of a partner.

The world would be less interesting if unicellular organisms in some ancient swamp had not teamed up, giving rise to multicellular species and, eventually, to us. But, at the same time as allowing new survival strategies (not to mention the emergence of civilization), the evolution of multicellularity presented new difficulties: for example, cells in a single organism, and therefore derived from the same egg and sperm, had to choose for the first time between many possible fates.

We now know that most cells learn their fate from signals produced by other cells, and the hunt is on to identify components of these signalling systems. On pages 507 and 512 of this issue, Weiss, Frasch and co-workers¹ and Palmer and colleagues² describe experiments that unite a pair of hitherto enigmatic signalling proteins — a secreted protein, Jeb, and a cell-surface-located receptor, Alk. By showing that Jeb binds to and activates Alk, these papers provide new insight into development, and also illustrate the relationship between development and cancer: developmental signals are potent regulators of cell behaviour, and so can have disastrous effects if uncontrolled.

This story began in 1997, when unregulated activity of the anaplastic lymphoma kinase (Alk) protein was discovered to be the cause of a cancer known as anaplastic large-cell lymphoma^{3–5}. Alk belongs to a family

of enzymes called receptor tyrosine kinases, and is most closely related to the insulin receptor. So it was straightforward to predict — and it was subsequently confirmed — that, like other members of this enzyme family, Alk is a cell-surface receptor that activates several intracellular signal-transduction cascades, including a pathway that contains the so-called mitogen-activated protein (MAP) kinase. It was, however, much harder to determine which protein (or proteins) binds to Alk to activate these pathways.

Simultaneously, but in apparently unrelated experiments, fruitfly biologists were

trying to understand how the intricate patterning of the embryonic musculature develops. Scott and colleagues⁶ had identified an unusual secreted protein, Jelly belly (Jeb), which was required for the development of the visceral muscles — those that are under involuntary control, such as the muscles that move food through the gut. They found that Jeb was emitted from adjacent somatic muscles, and proposed that it was a signal by which somatic muscles could induce neighbouring mesoderm tissue to produce visceral muscle cells.

In a collaborative effort, the laboratories of Manfred Frasch and Joseph Weiss have pursued this model, as they now describe¹. They examined fruitfly embryos with mutations in the *jeb* gene, and discovered that Jeb is required specifically for the formation of visceral 'founder' cells in mesoderm. It is not, however, needed to produce the follower cells that later fuse with these founders. The previous discovery that the founder cells internalize extracellular Jeb⁶ suggested that they express a specific receptor for this secreted protein. And the fact that MAP kinase is activated in these same cells¹ hinted that Jeb could be a binding partner for a receptor tyrosine kinase, as such receptors are known to trigger MAP kinase.

In a different approach, Ruth Palmer's laboratory had been using fruitflies to study the normal function of Alk (although its pathological function in cancer was known, its normal role was not). They showed that Alk is expressed in⁷, and required for the formation of⁸, the early visceral mesoderm in flies.

This was the point at which the two projects intersected: Frasch and colleagues saw the published expression pattern of Alk⁷ and speculated that it could be their predicted receptor tyrosine kinase; Palmer and co-workers saw the similarity of the *jeb* mutants⁶ to their *alk* mutants, and guessed that Jeb could be the binding partner for Alk. They independently used a powerful combination of genetics and biochemistry to show that

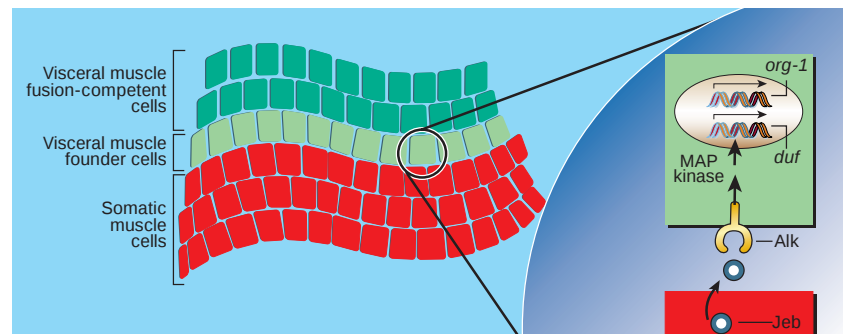


Figure 1 Making muscles in fruitflies. This model is based on the new findings^{1,2}. During the formation of visceral muscles (such as gut muscles), somatic muscles (red) secrete Jeb, while potential visceral founder cells (green) express the receptor Alk. Alk-expressing cells next to Jeb-secreting cells receive the signal, which activates the MAP kinase pathway. This induces expression of *duf* and *org-1*, thus inducing the Alk-expressing cells to become visceral muscle founder cells (pale green).

this is indeed the case^{1,2}. We now know that Jeb is secreted by somatic mesoderm cells and binds to Alk on the surface of adjacent mesoderm cells that are due to become visceral muscle (Fig. 1). This leads to Alk activating the MAP kinase pathway, via the Ras protein; this pathway in turn triggers the expression of the gene-transcription factors Duf and Org-1, which 'tell' the mesoderm cells to become visceral founder cells.

There are some further insights. The Frasch and Weiss groups¹ discovered that Jeb–Alk signalling simultaneously inhibits a transcription factor that is required for somatic muscle 'determination'. In this way, Jeb–Alk signalling definitively determines that cells become visceral, not somatic, muscles. Palmer's group² showed that internalization of the Jeb–Alk complex depends on the receptor's kinase activity, suggesting that the internalization and subsequent degradation of the signal might itself be regulated by signalling, possibly in some kind of feedback loop. Both papers agree that when cells are prevented from becoming visceral founders as usual, they abnormally fuse with somatic muscles.

These papers extend our understanding of how cell-to-cell signalling controls cell-fate decisions. They also marry two signalling proteins in search of partners. It will, of course, be a high priority to discover whether mammalian Jeb-like proteins also bind to Alk, and whether their role in muscle development is conserved. The fruitfly is a powerful model for mammalian biology, but

it cannot be assumed that any given system is identical. Indeed, two other proteins — pleiotrophin and midkine — have been shown to activate mammalian Alk^{9,10}, so the relative roles of these different potential binding partners must be assessed. But genetics has now definitively established that, in fruitflies at least, Jeb is a physiologically significant partner for Alk.

Finally, although one of the obvious attractions of these papers is the discovery of a normal function for a protein that is implicated in cancer, it is important to remember that the relationship between the developmental and tumour-causing functions of signalling proteins can be quite indirect. The cancer-inducing form of Alk is misexpressed and does not require a partner to activate it, and has therefore escaped from normal physiological control. It has run amok, and the damage it causes may be unrelated to its normal function. ■

Matthew Freeman is in the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

e-mail: MF1@mrc-lmb.cam.ac.uk

1. Lee, H.-H., Norris, A., Weiss, J. B. & Frasch, M. *Nature* **425**, 507–512 (2003).
2. Englund, C. et al. *Nature* **425**, 512–516 (2003).
3. Morris, S. W. et al. *Oncogene* **14**, 2175–2188 (1997).
4. Iwahara, T. et al. *Oncogene* **14**, 439–449 (1997).
5. Duyster, J., Bai, R. Y. & Morris, S. W. *Oncogene* **20**, 5623–5637 (2001).
6. Weiss, J. B., Suyama, K. L., Lee, H. H. & Scott, M. P. *Cell* **107**, 387–398 (2001).
7. Loren, C. E. et al. *Genes Cells* **6**, 531–544 (2001).
8. Loren, C. E. et al. *EMBO Rep.* **4**, 781–786 (2003).
9. Stoica, G. E. et al. *J. Biol. Chem.* **277**, 35990–35998 (2002).
10. Stoica, G. E. et al. *J. Biol. Chem.* **276**, 16772–16779 (2001).

Cell biology

The hippo hypothesis

Michael E. Rothenberg and Yuh-Nung Jan

The perfection of a fly's eye and the chaotic nature of tumours provide eloquent examples of the need to coordinate cell death and proliferation. The intricacies of the underlying mechanism are now being uncovered.

Just as the number of working parts in a machine is crucial, so too is the number of cells in a tissue. This means that the creation of new cells (by proliferation) and the elimination of excess ones (by programmed cell death, or 'apoptosis') must be tightly coordinated. Although many genes have been shown to regulate either proliferation or apoptosis, little is known about how the two are coupled. Recently, however, geneticists have begun to uncover some of the genes involved in this coordination, and, writing in *Cell*, Wu et al.¹ and Harvey et al.² describe another such gene, *hippo*. Their findings significantly advance our understanding of this fundamental problem in organ development and cancer biology.

Both groups^{1,2} looked at the imaginal

discs of fruitflies. Imaginal discs are packets of cellular monolayers (hence the name disc) that are set aside during larval development, and differentiate during metamorphosis (the pupal stage) to give rise to most of the tissues of the adult fly. In the case of the eye, on which the two groups focused, undifferentiated retinal imaginal disc cells proliferate rapidly throughout much of development. Later, waves of differentiation sweep across the eye disc to pattern the tissue (that is, to determine the fates and arrangements of the cells) and to halt cell division. This leads to the formation of a highly organized retina consisting of about 750 regularly spaced 'seeing units', called ommatidia (Fig. 1a, overleaf). During this process, the imaginal disc has an excess of cells between the ommatidia. Most of these leftover



100 YEARS AGO

Die Schule der Chemie. Prof. Ostwald is an ingenious man; in his own language, the attribute might be expressed by the adjective "schlau." Having, as he tells us in his preface, published volumes of the greatest importance, and of the widest range, on physical chemistry for the use of investigators in the domains of chemistry and physics... he now makes an attempt in this very elementary work to reach a larger public, and has written this most amusing book for the use of youngsters about ten to thirteen years of age. The plan adopted is to introduce by means of dialogue some chemical facts... Talking of the combustion of a candle and its disappearance, the pupil says, "But it really vanishes before my eyes." "Yes," says the teacher, "it becomes invisible. But can't it change into something invisible?" "There are no invisible things," says the pupil. "Oho!" replies the teacher. "No," says the pupil, "ghosts and goblins don't exist." "Even they are said to be sometimes visible," answers the teacher. "But can you see the air?" "Hum — no," says the pupil. "But the air is changed by burning. I don't see how." And so the formation of an invisible gas is brought out, and the method of determining its weight.

From *Nature* 1 October 1903.

50 YEARS AGO

The loofah of commerce, also known as the 'Luffa sponge', 'dishcloth' or 'dishrag gourd', 'snake gourd' and 'vegetable sponge', is the cleaned, dried inner fibrous tissue of the fruit or gourd of *Luffa* spp., which grow readily in most warm countries. It is a climbing annual plant, and has been grown in Japan from olden times. The plant and fruits resemble cucumbers; but the fruits are fatter and hang heavily on the vines, while the stems are more woody and stronger. In Japan, the plant blooms during June–July, and the flowers last well into the autumn... The best type of loofah for the commercial market is the one produced in Japan, where the soil and climate have proved well suited to their cultivation. An account of the loofah industry in Japan has been given by J. S. Ingram in *Colonial Plant and Animal Products*... The cultivation and possibilities of the plant in Colonial territories is also described.

From *Nature* 3 October 1953.

'interommatidial' cells (some 2,000) are later eliminated by a strictly controlled wave of apoptosis, leaving a thin layer. The end result is a retina so perfectly organized that it has been called a "neurocrystalline lattice"³. Its perfection provides a sensitive assay for probing the link between proliferation and apoptosis.

Previous work had already indicated that cell proliferation and death in the developing retina are tightly coupled. First, inducing apoptosis early on in a group of retinal precursors — a manipulation that should dramatically reduce the size of the eye if cell death and proliferation were completely independent and inflexible — does not result in a much smaller eye⁴. Second, driving extra proliferation causes the imaginal disc to grow too large only if cell death is simultaneously inhibited⁵. Third, several genes, including *salvador* (also known as *shar-pei*)⁶, have been shown to coordinately regulate both proliferation and apoptosis^{7,8}.

This coupling of proliferation and apoptosis seems to be a general rule in multicellular organisms, and the dysregulation of these processes can lead to cancer. Indeed, tumour development is thought to require both an increase in proliferation and a decrease in cell death⁹. Thus, a gene that promotes both apoptosis and exit from the cell-division cycle should be a tumour-suppressor gene. In fact, *salvador* is such a gene, and its human relative *hWW45* has been implicated in several cancers⁸.

Quite how *salvador* and other such genes couple cell proliferation and death was unknown. But our understanding of this is now advanced by the characterization of *hippo*^{1,2}, a gene that functions with *salvador* and another tumour suppressor called *warts* (also known as *lats*)¹⁰. Wu *et al.*¹ and Harvey *et al.*² embarked on genetic screens to identify genes that negatively regulate tissue growth. They found *hippo*, a gene that belongs to a family of kinases — enzymes that add phosphate groups to other proteins. The Hippo protein is 60% identical to the human kinase Mst2, a close relative of which, Mst1, has been implicated in apoptosis¹¹.

The authors then found^{1,2} that tissues lacking *hippo* contain too many cells. Within those mutant tissues, cell size and many aspects of patterning and cell-fate determination are remarkably normal — but the number of cells is dramatically increased. For instance, in the mutant eye, early patterning is not affected and photoreceptor cells differentiate and assemble into normal ommatidia. Yet there are many more interommatidial cells than normal, resulting in overgrown and folded eyes (Fig. 1b).

What causes the increase in cell number? First, there is increased proliferation. After the wave of differentiation has swept across the imaginal disc, normal cells stop dividing (or divide only once more). But cells lacking

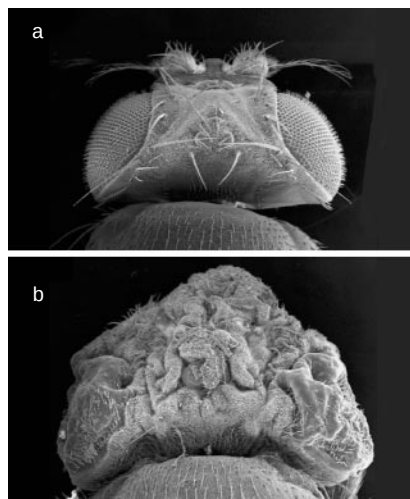


Figure 1 Eyes right, and wrong. a, The head of a normal fruitfly, showing its perfectly organized eyes (left and right). b, The new papers^{1,2} show that mutations in the *hippo* gene increase cell proliferation and decrease cell death, resulting in vastly overgrown tissues, including eyes. (Reproduced with permission from ref. 1.)

hippo continue to cycle for some time. At least one reason is that these cells upregulate Cyclin E, a limiting factor for entry into the DNA-replication phase of proliferation in imaginal disc cells^{5,12}. Both groups agree that transcription of the *Cyclin E* gene is increased, although Harvey *et al.* suggest that a post-transcriptional mechanism might also contribute. So why are these extra cells not simply pruned away by apoptosis? The reason is that *hippo* is also required for apoptosis. Cells lacking *hippo* have increased levels of a key inhibitor of apoptosis, DIAP1, and are strongly resistant to treatments that should induce apoptosis.

The defects produced by a lack of *hippo* are remarkably similar to those produced when *warts* or *salvador* is missing, and indeed Wu *et al.* observed strong, specific genetic interactions between the three genes. These experiments, as well as other studies described in the papers, show that Hippo can associate with and phosphorylate Salvador, and that this interaction helps Hippo to phosphorylate Warts, another kinase. This suggests a model in which Salvador functions as an adaptor protein that brings Hippo and Warts together in a ternary complex. Presumably, phosphorylated Warts — as the major (but not the only) output of this complex — then goes on to regulate Cyclin E, DIAP1 and other downstream effectors. This work has thus begun to define a genetic pathway for the coordinated regulation of cell-cycle exit and apoptosis.

So what remains to be discovered? First, there is the question of mechanism. What is the exact nature of the physical, genetic and biochemical interactions between Hippo, Salvador and Warts? What is the significance of the phosphorylation of Warts by Hippo?

And when, where and how does phosphorylation occur in this pathway *in vivo*? Also, given that any pathway that includes kinases also has regulatory phosphatases (proteins that remove phosphate groups), are phosphatases also involved in coordinating proliferation and death?

Second, there is the question of downstream targets. Tissues lacking *salvador*, *hippo* or *warts* show reduced cell death and more proliferation — but the cells also divide faster, yet remain of normal size. So this pathway also regulates cell growth (mass accumulation), and downstream targets in addition to DIAP1 and Cyclin E must exist. Might Hippo, as a member of the Ste20 family of kinases¹³, be connected to signalling modules that contain mitogen-activated protein kinases — modules that regulate cell proliferation, death and growth elsewhere?

Third, there is the question of upstream control. How do cells receive and process the signals that work through the Hippo–Salvador–Warts pathway to control proliferation, death and growth? As all three proteins are ubiquitously expressed, it is likely that their activity (or the ability of cells to respond to them) is regulated, rather than their expression. Perhaps, as in other pathways involving Ste20-family kinases, regulatory enzymes such as members of the Ras family will be found upstream.

Finally, as a first step towards testing whether Hippo's role in coupling proliferation to apoptosis is evolutionarily conserved, Wu *et al.*¹ investigated whether the human Mst2 protein could restore the status quo to fruitflies that lack Hippo. They found that it could, raising the exciting possibility that Mst1 and Mst2 are tumour suppressors. This question could be investigated in several ways: by knocking out the genes in mice; by analysing them in tumour cell lines; and by searching for human families with heritable cancer syndromes that are caused by mutations in the genes for Mst1 or Mst2. ■

Michael E. Rothenberg and Yuh-Nung Jan are in the Departments of Physiology and Biochemistry, Howard Hughes Medical Institute, University of California, San Francisco, California 94143-0725, USA. e-mail: ynjn@itsa.ucsf.edu

1. Wu, S., Huang, J., Dong, J. & Pan, D. *Cell* **114**, 445–456 (2003).
2. Harvey, K. F., Pfeiffer, C. M. & Hariharan, I. K. *Cell* **114**, 457–467 (2003).
3. Ready, D. F., Hanson, T. E. & Benzer, S. *Dev. Biol.* **53**, 217–240 (1976).
4. Stowers, R. S. & Schwarz, T. L. *Genetics* **152**, 1631–1639 (1999).
5. Neufeld, T. P., de la Cruz, A. F., Johnston, L. A. & Edgar, B. A. *Cell* **93**, 1183–1193 (1998).
6. Kango-Singh, M. *et al.* *Development* **129**, 5719–5730 (2002).
7. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. *Cell* **113**, 25–36 (2003).
8. Tapon, N. *et al.* *Cell* **110**, 467–478 (2002).
9. Green, D. R. & Evan, G. I. *Cancer Cell* **1**, 19–30 (2002).
10. Xu, T., Wang, W., Zhang, S., Stewart, R. A. & Yu, W. *Development* **121**, 1053–1063 (1995).
11. Graves, J. D. *et al.* *EMBO J.* **17**, 2224–2234 (1998).
12. Richardson, H. *et al.* *Development* **121**, 3371–3379 (1995).
13. Dan, I., Watanabe, N. M. & Kusumi, A. *Trends Cell Biol.* **11**, 220–230 (2001).

Obituary

Patricia S. Goldman-Rakic (1937–2003)

It was a snowy winter day in 1998 and I had just heard Patricia Goldman-Rakic talk about her recent studies of the cerebral cortex and its abnormal functioning in schizophrenia. Driving home together from Massachusetts, the slow traffic gave us ample time to delve into the implications of her findings. Goldman-Rakic's work addressed one of the last frontiers of human knowledge: the neural mechanisms in the brain that control thought and planning, reasoning and attention. She used a uniquely wide array of experimental tools, in non-human primates and in humans, to explore these mechanisms under normal and pathological conditions. Yet this only begins to explain the enormous loss to the field caused by her untimely death in a car accident on 31 July 2003.

Goldman-Rakic's work focused on the brain's frontal lobes, in particular the prefrontal cortex. From an evolutionary point of view, this is the most recently developed part of the brain, and is greatly advanced in primates compared with other mammals. Still, not much was known about it in 1965, when Goldman-Rakic started working on the region as a junior research fellow at the National Institute of Mental Health (NIMH) in Bethesda. She had recently completed her PhD at the University of California, Los Angeles, where she studied rodent models of stress.

Over the next ten years, Goldman-Rakic and her colleagues, along with other groups, created small lesions in the brains of adult rhesus monkeys and other non-human primates, documenting for the first time that the prefrontal cortex is essential for a form of cognition termed working memory or executive function. This refers to the ability of an individual to keep something 'in mind' while attending to other tasks. Working memory was measured with 'delayed-response' tasks; these come in many forms, but the basic principle is that a monkey is required to maintain a mental representation of something, such as where a food reward is located, which guides its subsequent responses to obtain this reward.

In the 1970s, by which time Goldman-Rakic was chief of the Section on Developmental Neurobiology at NIMH, she and her colleagues had extended this work to show that damage to the monkey prefrontal cortex early in life causes cognitive abnormalities that become noticeable during adolescence. At about the same time, she demonstrated a crucial



Neuroscientist who revolutionized the study of higher brain function

role for the neurotransmitter dopamine in regulating prefrontal cortical function. Both findings related to schizophrenia, a common and debilitating syndrome that was at the time defined by 'positive symptoms' (hallucinations and delusions) and 'negative symptoms' (blunted emotions). As with Goldman-Rakic's monkeys, schizophrenia develops in adolescence and early adulthood. Moreover, dopamine was known to influence the symptoms of schizophrenia, because all anti-schizophrenic medications were (and still are) molecules that block dopamine receptors on neurons.

In 1979 Goldman-Rakic moved her laboratory to Yale University, where she joined the newly created Section of Neuroanatomy. The years that ensued brought remarkable productivity, as her research evolved in several new directions. She characterized the three-dimensional structure of individual prefrontal cortical nerve cells and the several thousand connections (synapses) that each nerve cell forms with its inputs. She identified parallel streams of sensory information that feed into the prefrontal cortex. She and her co-workers also defined exactly where on these nerve cells different types of receptor for dopamine and other neurotransmitters are located.

The Goldman-Rakic laboratory also began to record the electrical activity of individual prefrontal cortical neurons while monkeys were performing delayed-response tasks. They discovered nerve cells that respond to visual and spatial cues

involved in working memory, but not to other types of sensory stimulus. This work supported the view that the prefrontal cortex is organized according to sensory modality, and defined several subregions of the prefrontal cortex into which these modalities are segregated. In addition, Goldman-Rakic discovered how dopamine affects the function of individual neurons and correlated the different effects of dopamine with particular behavioural actions, which ranged from improved attention to impaired performance in working-memory tasks.

Goldman-Rakic applied this growing understanding of the prefrontal cortex in monkeys to studies of schizophrenia. Her laboratory carefully defined cellular abnormalities in autopsy specimens taken from the prefrontal cortex of people with schizophrenia. By the use of brain imaging, her group began to define areas of the prefrontal cortex that function abnormally in schizophrenia. Although the exact nature of these abnormalities in the prefrontal cortex is still incompletely understood, Goldman-Rakic's work helped to spark new ideas about where in the brain to search, and what types of cellular abnormalities to look for, in schizophrenia. Indeed, the growing evidence for a role of the prefrontal cortex in schizophrenia drove the field — relatively recently — to recognize cognitive abnormalities, particularly in working-memory processes, as a third major clinical feature of this syndrome.

Translational — bench-to-clinic — research in neuroscience is as difficult as it is necessary, yet Goldman-Rakic succeeded brilliantly. She tackled the complex nature of working memory at the neural, cellular and molecular levels. Moreover, her findings have led to a fundamental revisiting of the effects of dopamine on cortical function, and hence to efforts to develop new drugs for treating schizophrenia.

The depth and breadth of Goldman-Rakic's contributions to the study of the prefrontal cortex and schizophrenia led to many honours. In a field still dominated by men, she was one of all too few senior women scientists. Her death deprives the field of still more discoveries and insights. But her work defines the elegant process of scientific inquiry that will be needed to solve the riddle of thinking, and its derangement in schizophrenia. Eric J. Nestler
Eric J. Nestler is in the Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9070, USA.
e-mail: eric.nestler@utsouthwestern.edu

Cancer

Further perils of passive smoking

Cancer Cell **4**, 191–196 (2003)

Passive smoking is thought to cause lung cancer, but this might be only one of its effects. Research by Bo-qing Zhu *et al.* hints that second-hand smoke could also speed tumour growth by prompting new blood vessels to form.

The team exposed mice that had been injected with lung cancer cells to second-hand smoke or clean air for 17 days. Fume-breathing animals developed tumours that were about five times larger than those in control animals. Blood levels of vascular endothelial growth factor (VEGF) — a molecule that stimulates blood-vessel growth — were doubled, and the tumours developed elaborate blood supplies.

When the smoke-breathing mice were treated with mecamylamine, which stops nicotine from binding to cells, tumour growth was reduced by two-thirds. This suggests that the nicotine found in second-hand smoke stimulates blood-vessel and tumour growth. Drugs that block the receptors for nicotine or VEGF may help to slow tumour growth, the team speculate.

Helen R. Pilcher

Plate tectonics

America claims the Far East

Geophys. Res. Lett. **30**, 1924–1927 (2003)

Part of Siberia belongs to America. That, at least, would be the case if national boundaries were defined tectonically rather than politically and geographically. G. M. Steblov *et al.* — an even-handed collaboration of Russian and American

geoscientists — have mapped out the boundaries of the tectonic plates in the eastern corner of Siberia, and find that everything east of the Cherskiy mountain range, lying roughly along latitude 160° E, is part of the North American plate (see picture). West of this, the land belongs to the Eurasian plate.

This division has been long suspected, even taken for granted. But the researchers have now been able to confirm it, thanks to data on plate motions obtained over a 4–6-year period using the Global Positioning System. The relative motions of several Siberian monitoring stations indicate which plate they are situated on. The results do not at this stage resolve the question of whether there are two small microplates — the Okhotsk and Amurian — at this boundary. But if the Okhotsk microplate does not exist, the North American plate snips off the northernmost tip of Hokkaido in Japan as well.

Philip Ball

Physiology

Fat body sizes fly

Cell **114**, 739–749 (2003)

Apart from storing proteins, carbohydrates and fat, the so-called fat body in the fruitfly *Drosophila* also secretes hormones — not unlike our liver. Julien Colombani and colleagues claim that it also senses nutrient availability and orchestrates fly growth.

In a screen for genes that modify growth, Colombani *et al.* identified *slimfast*, a gene encoding a protein that transports amino acids into cells. Lowering the amount of *slimfast* in *Drosophila*'s fat-body cells mimicked nutrient deprivation in only these cells, yet it triggered a widespread response: the development of fly larvae was delayed and, even though food intake was normal, the flies were born small and lean, with little wings.

Cells can limit their growth by turning off enzymes such as PI3K. And when flies are fed amino-acid-free food, this is exactly what happens. But individual cells grown in the absence of amino acids don't switch off PI3K, suggesting that amino-acid availability is not sensed at the cellular level, but by some central sensor. Could that sensor be the fat body?

Indeed, Colombani *et al.* show that an amino-acid shortage in the fat body suffices to suppress PI3K activity in the fly's peripheral tissues. They propose that the fly's fat body is its nutrient sensor — which, presumably by the secretion of hormones, adjusts fly size according to food availability.

Marie-Thérèse Heemels

Astrophysics

Lost lithium

Preprint at <<http://arXiv.org/astro-ph/0309480>> (2003)

Light elements such as deuterium, helium and lithium were created in the furnace of the early Universe. Since then, they have been slowly consumed in stars. Only in old, slowly simmering stars and tenuous gas clouds has the relative quantity apparently remained more or less unchanged.

But old stars have lost more lithium than astronomers thought, according to Alain Coc and colleagues. Using the latest tally of atoms in the Universe, calculated from observations of the cosmic microwave background radiation, and updated estimates of nuclear reaction rates inside stars, Coc *et al.* predicted how much lithium they should see. Only a third of this amount is evident in old stars.

In contrast, deuterium — which is far more easily destroyed — has not changed much in abundance over the Universe's lifespan. The authors suggest that the estimated nuclear reaction rates for lithium may have to be revised.

Joanne Baker

Cell biology

It's the old story of yeast

Science **301**, 1908–1911 (2003)

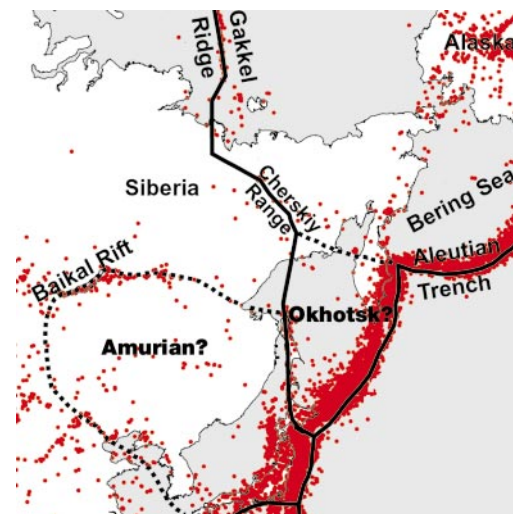
The incidence of cancer increases with age. And genomic instability is thought to contribute to tumour development. Are the two connected? From studies on baker's yeast, Michael A. McMurray and Daniel E. Gottschling have evidence that they may be.

A common measure of yeast lifespan is the number of daughter cells a mother cell produces before dying. McMurray and Gottschling isolated successive daughter cells over the entire lifespan of a mother, and looked for the presence of an indicator of genomic instability in certain marker genes. That indicator — 'loss of heterozygosity' — was more common in daughter cells from older mother cells than in those from younger ones, apparently because of an increased rate of chromosome breaks. Once loss of heterozygosity was detected, it occurred at greater frequencies in subsequent progeny. So there seems to be a switch, thrown late in life, that makes the yeast genome more unstable, and it is on a clock that is distinct from the clock that determines yeast lifespan.

The authors speculate that ageing mother cells may accumulate damaged proteins that are less able to repair damaged DNA. But whether similar mechanisms contribute to the age-related increased occurrence of human cancers is, of course, another matter.

Barbara Marte

AGU



Tectonic takeover: the North American plate includes part of Siberia, and might also claim the tip of Japan. Red dots indicate the incidence of earthquakes.

Captivity effects on wide-ranging carnivores

Animals that roam over a large territory in the wild do not take kindly to being confined.

Some species — ring-tailed lemurs and snow leopards, for example — apparently thrive in captivity, whereas others, such as Asian elephants and polar bears, are prone to problems that include poor health, repetitive stereotypic behaviour and breeding difficulties. Here we investigate this previously unexplained variation in captive animals' welfare by focusing on caged carnivores, and show that it stems from constraints imposed on the natural behaviour of susceptible animals, with wide-ranging lifestyles in the wild predicting stereotypy and the extent of infant mortality in captivity. Our findings indicate that the keeping of naturally wide-ranging carnivores should be either fundamentally improved or phased out.

Preventing natural behaviour patterns in animals can give rise to stress and frustration^{1,2}, and impair the development of brain regions that are involved in behavioural sequencing, thereby reducing the animal's ability to behave flexibly and appropriately^{3,4}. To investigate whether the observed variation in the welfare of different species could arise from a differential impact of captivity on their natural behaviour, we calculated the mean frequency of stereotypic pacing⁵ by 35 species of caged carnivore. We focused on pacing because it is the most prevalent stereotypy among carnivores (97% of reported stereotypies⁵) and also to avoid comparability problems raised by pooling different forms of stereotypy (such as swaying and head-nodding). We also quantified infant mortality in captivity, which is often due to poor maternal care⁶.

As an animal's natural ranging and foraging activities are particularly constrained by captivity⁷, we obtained all available field data on median home-range size, daily travel distance, time spent in general activity, time spent foraging, and reliance on hunting. We also quantified minimum home-range sizes and daily travel distances, as these can be orders of magnitude smaller when food is abundant⁸. Relationships between wild and captive variables were tested by using one-tailed regressions.

Body-weight effects were investigated in analyses involving range size⁹; phylogenetic effects were controlled where necessary (and in all analyses involving body mass) by comparative analysis of independent contrasts^{10,11}. Our inferences about welfare took into account natural infant-mortality rates, and the amount of normal activity and total stereotypy in captivity; we also considered feeding regimes, and the size and complexity of enclosures, to check that relationships between wild and captive variables were

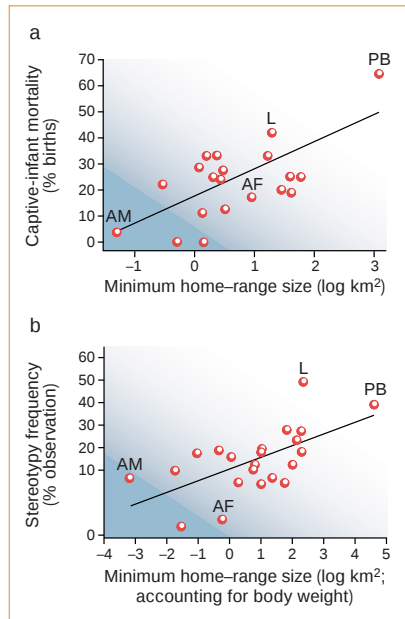


Figure 1 Natural ranging behaviour and welfare of species from the order Carnivora in captivity. **a**, Carnivores' minimum home-range sizes in the wild predict captive infant mortality ($F_{1,19} = 12.60$, $P = 0.001$). **b**, Together with body weight (see text), minimum home-range size also predicts stereotypic pacing in captivity ($F_{2,19} = 4.79$, $P = 0.011$; controlling for phylogeny: $F_{2,17} = 3.11$, $P = 0.036$). On these cross-species plots, a few species from a range of families and with varying relation to the regression line are highlighted: AF, Arctic fox (*Alopex lagopus*); PB, polar bear (*Ursus maritimus*); AM, American mink (*Mustela vison*); L, lion (*Panthera leo*). Values on the x-axes differ because fitted values are used in **b** that incorporate body weight; the y-axis in **b** shows data back-transformed from an arc-sine transformation.

not by-products of variation in husbandry. Degrees of freedom varied in subsequent analyses owing to missing data.

Natural home-range size (HR) predicted captive-infant mortality (median HR: $F_{1,21} = 6.04$, $P = 0.012$; minimum HR, see Fig. 1a). Controlling for body weight did not alter this relationship (median HR: $F_{1,20} = 4.35$, $P = 0.025$; controlling for phylogeny: $F_{1,16} = 20.46$, $P = 0.0001$; minimum HR: $F_{1,18} = 9.29$, $P = 0.004$; controlling for phylogeny: $F_{1,18} = 16.94$, $P = 0.001$). Minimum, but not median, daily distances travelled (DDT) gave similar results ($F_{1,18} = 3.99$, $P = 0.03$). These effects seem to be specific to captive animals: wild and captive infant-mortality rates did not covary ($F_{1,6} = 0.08$, not significant) and infant mortality in the wild was unrelated to range size (for example, minimum HR: $F_{1,7} = 0.43$, n.s.).

Home-range size also predicted pacing (median HR: $F_{1,22} = 5.78$, $P = 0.013$; minimum HR: $F_{1,20} = 5.66$, $P = 0.014$). A positive

trend was evident with body weight ($F_{1,33} = 3.23$, $P = 0.081$; controlling for phylogeny: $F_{1,31} = 4.09$, $P = 0.052$). With both terms in a multiple regression, each lost its individual effect on pacing, but the overall adjusted r^2 value increased to 26.5%, from 6.2% (for body weight alone) and 18.2% (for minimum HR alone; Fig. 1b). Likewise, median, but not minimum, daily travel distances were positively correlated with pacing ($F_{1,18} = 9.80$, $P = 0.003$).

These results all held when total stereotypical behaviours (including non-pacing) were analysed⁵. Naturally wide-ranging animals did not, however, show more normal activity in captivity (for example, minimum HR: $F_{1,15} = 0.17$, n.s.; DDT: $F_{1,15} = 0.01$, n.s.), nor did they move around more overall within their enclosures (for example, minimum HR: $F_{1,18} = 0.61$, n.s.; DDT: $F_{1,18} = 0.10$, n.s.). Home-range size therefore still predicted pacing, even when controlling for the amount of total activity in captivity (for example, minimum HR: $F_{1,14} = 2.65$, $P = 0.019$). The degree of natural foraging and general activity, in contrast, did not predict captive stereotypy or infant mortality⁵ (for example, the level of natural activity versus pacing: $F_{1,18} = 3.53$, n.s.). Variations in husbandry did not account for any of these findings⁵.

Our results show, to our knowledge for the first time, that a particular lifestyle in the wild confers vulnerability to welfare problems in captivity. Our study also reveals species that are inherently likely to fare badly in zoos and similar establishments. Among the carnivores, naturally wide-ranging species show the most evidence of stress and/or psychological dysfunction in captivity^{3,4,12}, a finding that is a cause for concern, given the difficulties of conserving such species *in situ*¹³. Husbandry of these species in captivity is therefore in need of improvement, such as provision of extra space (a polar bear's typical enclosure size, for example, is about one-millionth of its minimum home-range size). Alternatively, zoos could stop housing wide-ranging carnivores and concentrate instead on species that respond better to being kept in captivity.

Ros Clubb, Georgia Mason

Animal Behaviour Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

e-mail: georgia.mason@zoology.ox.ac.uk

1. Mason, G., Cooper, J. & Clarendon, C. *Nature* **410**, 35–36 (2001).
2. Dawkins, M. S. *Appl. Anim. Behav. Sci.* **20**, 209–225 (1988).
3. Robbins, T. W., Jones, G. H. & Wilkinson, L. S. *J. Psychopharmacol.* **10**, 39–47 (1996).
4. Lewis, M. H., Gluck, J. P., Bodfish, J. W., Beauchamp, A. J. &

- Mailman, R. B. in *Stereotyped Movements* (eds Sprague, R. L. & Newell, K. M.) 37–67 (Am. Psychol. Assoc., Washington DC, 1996).
5. Clubb, R. E. *The Roles of Foraging Niche, Rearing Conditions and Current Husbandry on the Development of Stereotypes in Carnivores* (Thesis, Univ. Oxford, 2001).
6. Meier, J. in *Zoo and Wild Animal Medicine* (ed. Fowler, M. E.) 842–851 (Saunders, Philadelphia, 1986).
7. Shepherdson, D. J., Mellen, J. D. & Hutchins, M. (eds) *Second Nature: Environmental Enrichment for Captive Animals* (Smithsonian Inst. Press, Washington DC, 1998).
8. Gittleman, J. L. & Harvey, P. H. *Behav. Ecol. Sociobiol.* **10**, 57–63 (1982).
9. Gittleman, J. L. *J. Mamm.* **67**, 23–36 (1986).
10. Purvis, A. & Rambaut, A. *Comp. Appl. Biosci.* **11**, 247–251 (1995).
11. Bininda-Emonds, O. R. P., Gittleman, J. L. & Purvis, A. *Biol. Sci.* **74**, 143–175 (1999).
12. Bahr, N. I., Pryce, C. R., Döbeli, M. & Martin, R. D. *Physiol. Behav.* **64**, 429–437 (1998).
13. Woodroffe, R. & Ginsberg, J. R. *Science* **280**, 2126–2128 (1998).

Competing financial interests: declared (see online version).

Explosives

A microsensor for trinitrotoluene vapour

Sensing devices designed to detect explosive vapours are bulky, expensive and in need of technological improvement — dogs remain the most effective detectors¹ in the fight against terrorism and in the removal of land-mines^{2,3}. Here we demonstrate the deflagration of trinitrotoluene (TNT) in a small localized explosion on an uncoated piezoresistive microcantilever. This explosive-vapour sensor, which has a detection capability that is comparable to that of a dog, should enable extremely sensitive, miniature detection devices to be used on a large scale.

Microcantilevers with specific coatings can be used for chemical detection^{4,5}, but TNT molecules, which are intrinsically ‘sticky’, readily adhere to uncoated cantilever surfaces⁶. When the source of TNT is removed, the TNT molecules slowly desorb. We found that applying a voltage pulse to an integrated piezoresistor during this desorption process causes the cantilever temperature to rise beyond the deflagration point of TNT, resulting in a miniature deflagration. Our cantilevers did not degrade after hundreds of pulses (10–25 volts) and deflagrations.

We monitored deflagration by capturing magnified high-speed video images, by measuring the deflection due to released heat using an optical beam-bounce technique that is used in atomic-force microscopy, and by measuring resonance-frequency shifts and heat added with a piezoelectric/piezoresistive cantilever⁷ operated in self-sensing mode by means of an a.c. bridge circuit⁸. The circuit’s output increases with mass loading of adsorbed TNT and increasing temperature; it decreases with mass desorption and decreasing temperature.

A five-event TNT-detection test, using the self-sensing platform described, is illustrated

in Fig. 1. During event 1, before loading with TNT, a reference voltage pulse (25 volts) is applied to the piezoresistive heater, causing a temporary upward spike in circuit output that is due to heating, TNT loading (event 2) causes a gradual upward shift in sensor output, which then gradually decreases when the TNT begins to desorb from the cantilever (event 3). The second pulse (5 volts) during desorption does not raise the cantilever temperature sufficiently for deflagration (event 4). The third pulse (25 volts) causes deflagration, as shown by a visible smoke plume, and a dramatic mass decrease, which is verified by a reduction in circuit output (event 5) that overwhelms the upward thermal signal evident in event 1. Post-deflagration reference pulses of 25 volts resulted in spikes similar to the one seen in event 1.

The occurrence of deflagration was inferred from three consistent observations. First, the cantilever returns to its pre-test resonance frequency after deflagration, suggesting that all of the adsorbed material has been lost. Second, a specific voltage (corresponding to a threshold, or deflagration-point, temperature) is necessary to cause deflagration. Third, the measurement of heat added to the cantilever during deflagration shows that the reaction is exothermic, ruling out other possible reactions such as melting, vaporization or decomposition.

Our method currently detects the deflagration of as little as 70 picograms (1.9×10^{11} molecules) of TNT (calculated from the shift in cantilever resonance). This limit of detection is the same as that of an improved version⁹ of the ion-mobility mass-spectrometry technology now used for airport security. Calculations show that the detection limit could be improved by up to three orders of magnitude by using optimized cantilevers.

An explosive-detection technique based on deflagration is advantageous, because deflagration is a characteristic property of these substances. Other potentially interfering but non-explosive substances tested — including water, acetone, ethyl alcohol and gasoline — all desorbed in our system before the voltage pulse was applied. In the absence of a voltage pulse, desorption of nanogram quantities of deposited TNT takes tens of minutes; comparable amounts of water or alcohol desorb in seconds.

Other explosive molecules can be detected by this method and differentiated by their compound-specific deflagration points and desorption times (results not shown). Our technique could also be used in conjunction with coated-cantilever chemical-sensing arrays⁵, in which the arrays serve as an initial indicator, and positive readings are verified by deflagration on an uncoated cantilever.

L. A. Pinnaduwa*†, **A. Gehl***†, **D. L. Hedden***†, **G. Muralidharan***, **T. Thundat***†, **R. T. Lareau‡**, **T. Sulchek§**, **L. Manning§**, **B. Rogers§**, **M. Jones§**, **J. D. Adams§**

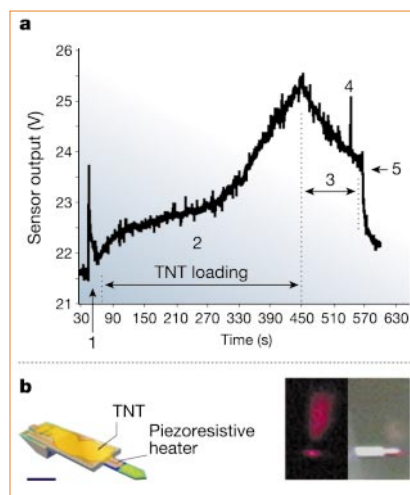


Figure 1 Deflagration and detection of trinitrotoluene (TNT). **a**, Sensor output plotted against time. During event 1 (numbered arrow) of the test, a 25-volt reference pulse is sent to the piezoresistor before TNT loading begins; sensor output returns quickly to almost its pre-pulse value. In event 2, TNT is gradually loaded onto the cantilever and the resonance frequency decreases, causing the sensor output to increase. After loading, the TNT begins to desorb (event 3) and a 5-volt pulse (event 4) is insufficient to heat the explosive on the cantilever to its deflagration point. However, a 25-volt pulse is sufficient to cause deflagration (event 5). **b**, Left, diagram of the microcantilever, showing the integrated piezoelectric sensor, piezoresistive heater, and TNT loaded on the cantilever surface. Scale bar, 100 μm . Right, deflagration is confirmed in magnified high-speed video images; the smoke plume is visible, particularly when illuminated by red laser light.

*Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

†Department of Physics, University of Tennessee, Knoxville, Tennessee 37996, USA

‡Transportation Security Administration, US Department of Homeland Security, Atlantic City, New Jersey 08405, USA

§Department of Mechanical Engineering and the Nevada Ventures Nanoscience Program, University of Nevada, Reno, Nevada 89557, USA

e-mail: jdadams@unr.edu

1. Furton, K. G. & Myers, L. J. *Talanta* **54**, 487–500 (2001).
2. Colton, R. J. & Russell, J. N. *Science* **299**, 1324–1325 (2003).
3. Czarnik, A. W. *Nature* **394**, 417–418 (1998).
4. Thundat, T., Chen, G. Y., Warmack, R. J., Allison, D. P. & Wachter, E. A. *Anal. Chem.* **67**, 519–521 (1995).
5. Lang, H. et al. *Appl. Phys. Lett.* **72**, 383 (1998).
6. Muralidharan, G. et al. *Ultramicroscopy* **97**, 433–439 (2003).
7. Minne, S. C., Manalis, S. R. & Quate, C. F. *Appl. Phys. Lett.* **67**, 3918–3920 (1995).
8. Rogers, B. et al. *Rev. Sci. Instruments* (in the press).
9. Khayamian, T., Tabrizchi, M. & Jafari, M. T. *Talanta* **59**, 327–333 (2003).

Competing financial interests: declared none.

brief communications is intended to provide a forum for brief, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers (**communications arising**). Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com/nature).

Binary switches and modification cassettes in histone biology and beyond

Wolfgang Fischle, Yanming Wang & C. David Allis

Laboratory of Chromatin Biology, The Rockefeller University, New York, New York 10021, USA

An immense number of post-translational modifications on histone proteins have been described and additional sites of modification are still being uncovered. Whereas many direct and indirect connections between certain histone modifications and distinct biological phenomena have now been established, concepts for comprehending the extreme density and variety of these covalent modifications are lacking. Here, we formally introduce localized 'binary switches' and 'modification cassettes' as new concepts in histone biology, elucidating mechanisms that might govern the biological readout of distinct modification patterns. Specifically, our hypotheses provide missing models for the dynamic readout of stable histone modifications and offer explanations for several long-standing questions embedded in the literature. Our ideas might also apply to non-histone proteins and are open to direct experimental examination.

Inside a eukaryotic cell nucleus, chromatin is far more than a static carrier of the genetic information encoded in DNA, as it actively mediates dynamic changes in gene function and expression¹. The chromatin polymer integrates a variety of endogenous and exogenous signals and functions as a biological relay station and signalling platform². Emerging evidence suggests that covalent post-translational modifications of histones—the main protein component of chromatin—play key roles in controlling the capacity of the genome to store, release and inherit biological information. Histone modifications can be highly reversible, such as histone (lysine) acetylation and histone (serine and threonine) phosphorylation, or more stable, such as histone (lysine and arginine) methylation^{3,4}. Hence, a wide range of histone- and chromatin-based regulatory options is available. These include rapid adjustments of gene expression in response to physiological and environmental stimuli as well as more permanent indexing systems, which could be involved in transmitting inheritable expression patterns from one cell generation to the next. Such inheritable changes in gene function and expression that don't involve changes in DNA sequence are commonly referred to as 'epigenetic'.

Fundamental cellular mechanisms uncovered in classical signalling pathways are manifested in the genetic and epigenetic regulatory circuits that control the post-translational modification of histones. For example, certain histone modifications (single sites of modification will be referred to as 'marks') recruit binding/effector proteins (referred to as 'modules')—a general biological principle initially established in phosphorylation-dependent signalling cascades⁵. Also, some histone modifications appear to act redundantly, a phenomenon that has been observed in several biological systems that rely on a multitude of singular post-translational modifications to achieve a robust signal⁶. However, other aspects of histone and chromatin biology appear to be more complex and are less clear. For example, whereas the binding of certain modules (that is, bromodomains and chromodomains) to histone marks is well documented, little is known about the release of such factors, especially from sites of modification that are presumed to be stable (that is, lysine-methylation). Particularly intriguing is the extreme high versatility and density of post-translational modifications found on histones. We believe that these features have evolved to sustain certain as yet unknown functions, an idea which has ultimately been expressed in the histone code hypothesis⁷. Here, we propose some general, but hypothetical, principles of how the combination of post-

translational modifications could be used to control effector-histone interactions, and mediate critical biological functions.

Dense clustering of histone modifications

A multitude of marks has been discovered on histone proteins³ and additional sites of previously unrecognized histone modification are still being unveiled, many by the recent introduction of mass spectrometry approaches to histone biology^{8,9}. Already, several short clusters of adjacent or closely-spaced modifiable residues are observed in all core histones and well studied 'hot spots' of clustered marks punctuate the tails of H3 and H4 (Fig. 1). Whereas *in vitro* experiments using purified enzymes imply that only a limited number of potential combinations of marks on single histones and possibly nucleosomes might exist¹⁰, initial mass spectrometric analysis of histones purified from various cellular sources point to more complex modification patterns *in vivo* (D. F. Hunt and C.D.A., unpublished observations). It seems unlikely that the multiple covalent marks found on histones have evolved independently. The extreme high density and versatility observed might instead serve valuable biological purposes—especially since most sites of post-translational modification are extremely conserved. If local clusters of marks were used in a combinatorial fashion, what mechanisms could direct their readout?

Local binary switches

Differential readout of distinct combinations of marks could, for example, be achieved by 'local switching' mechanisms such as the 'methyl/phos switch' proposed in Fig. 2. According to this model, phosphorylation of a site adjacent to (or nearby) a methyl mark that engages an effector module could lead to consecutive loss of binding to that factor. We hypothesize that the observed weak interaction of chromatin modules with their cognate marks could generally allow for rapid 'on-off' binding, which is then further regulated by modification of another neighbouring site in the 'off' state ($K_d \approx 10^{-4}$ – 10^{-6} M for both chromodomain/methyl-lysine and bromodomain/acetyl-lysine interactions). Such a kinetically controlled binding mechanism of effector modules is sustained by the recent demonstration that binding proteins of heterochromatic methyl marks have a fast exchange rate with their environment and are not statically bound to their target sites^{11,12}.

In particular, 'binary switches' could add an important dynamic nature to the readout of marks that themselves have very low turnover rates. For example, it is currently not understood if and how methyl marks in histones are reversed, and histone-

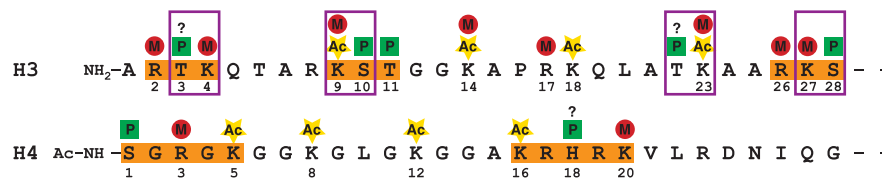


Figure 1 Dense clustering of histone marks. Histone modifications clustered in the N-terminal tails of H3 and H4. The sequences of the first 28 residues of human H3 and H4 are denoted, and sites of post-translational modifications are indicated. Note that lysine residues can be mono-, di- and tri-methylated and arginine residues can be mono- and di-methylated (symmetric or asymmetric)¹⁰. For clarity, these different methylation events are collectively represented as methyl marks. However, it is now clear that the additional complexity realized in different methylation levels is of great biological significance³⁹. Question marks accompany modifications that have been less

well documented. Methylation of Lys 23 has previously been reported in Alfalfa²⁵ and was verified by mass spectrometric analysis using H3 isolated from HeLa cells. Similarly, Thr 3 and Thr 22 have recently been identified as novel phospho-acceptors in the H3 tail (D.F. Hunt and C.D.A., unpublished observations). An older literature provides clues that His 18 might be an acid-labile phosphorylation site in H4 (ref. 34). Purple boxes indicate putative 'methyl/phos switch' sites; possible 'modification cassettes' in the tails of H3 and H4 are highlighted (see text for details). Ac, acetylation; M, methylation; P, phosphorylation.

demethylating enzymes are unknown¹³. Some histone methyl-lysine marks appear to be stably maintained over several cell generations¹⁴; yet, it is conceivable and likely that proteins bound to methyl marks are removed at some point in the cell cycle or during development of a complex organism. Therefore, we suggest that 'methyl/phos switching' could be a widespread mechanism regulating the binding and release of effector modules to more stable methyl marks. Nevertheless, we note that similar switches could also be operational for more labile marks (that is, 'acetyl/phos' or 'ubiquitin/phos' switches). In addition, the binding of yet to be discovered phospho-binding effector modules ('Y' in Fig. 2) could be regulated by nearby or adjacent 'off' switches.

H3 Lys 9/Ser 10 as a putative 'methyl/phos switch'

A site where a binary 'methyl/phos switch' might be operational is the Lys 9/Ser 10 region of H3. Methylation of Lys 9 by SET-type histone methyltransferases (HMT) like Su(Var)3-9 and G9a has been well documented and is associated with the establishment and maintenance of heterochromatic domains in many organisms⁴. The chromodomain of HP1 binds specifically to this mark and local HP1 recruitment is sufficient for mediating heterochromatin formation¹⁵ and accompanies gene silencing¹⁴. Genome-wide mitotic phosphorylation of H3 Ser 10 is catalysed by aurora B-type kinases, and several other enzymes mediate a more localized and targeted employment of phospho-Ser 10 in response to immediate-early gene signalling^{2,16}. Whereas initial reports using enzymatic *in vitro* assays pointed to a mutually exclusive existence of the methyl-Lys 9 and phospho-Ser 10 marks¹⁷, analysis of the *in vivo* modification pattern of H3 isolated from HeLa cells points to the coexistence of both marks on the same histone tail—especially during mitosis (D. F. Hunt and C.D.A., unpublished observations).

Structural examination of the chromodomain of HP1 bound to the H3 tail methylated on Lys 9 (refs 18, 19), allows the prediction that additional phosphorylation of Ser 10 could severely diminish the binding affinity of HP1 to its cognate mark (that is, release of 'X' in Fig. 2). Indeed, in *in vitro* assays chromodomains show almost complete loss of binding to dual modified peptides containing the methyl-Lys 9 and phospho-Ser 10 marks (W.F., S. A. Jacobs, S. Khorsanizadeh and C.D.A., unpublished observations). If the general concept depicted in Fig. 2 is correct, mitosis (or meiosis) or pathways of gene activation may drive the phosphorylation of the proposed 'methyl/phos switch', allowing for the release and potential clearing of 'negative' chromatin effectors like HP1 that repress transcription. This sequence of events could consecutively permit the docking of positive effectors that drive transcription (for example, HAT-containing complexes). In support, HP1 is partially liberated from interphase heterochromatin domains as cells enter mitosis^{20,21}, a cell cycle stage marked by a rapid, transient burst of H3 Ser 10 phosphorylation². Furthermore, HP1 displays an increased

mobility in T cells after receptor-driven kinase signalling is activated, an observation that is paralleled by a decrease of the immobile fraction of the protein¹¹.

Additional support for a Lys 9/Ser 10 'methyl/phos switch' is derived from the finding that type 1 protein phosphatase (PP1) acts as a mitotic phosphatase targeting phospho-Ser 10 as cells exit mitosis²². PP1 has independently been identified in genetic screens in *Drosophila* as Su(Var)3-6 (ref. 23). Genes of the Su(Var) family facilitate heterochromatic gene silencing as assayed by suppression of position effect variegation by mechanisms that have remained elusive for a long time. Our model of 'methyl/phos switches' makes a clear and testable prediction regarding the role of Su(Var)3-6 in the above silencing pathway: one role, if not the major role, of PP1 is to remove mitotic phosphates at Ser 10 in H3, thus allowing increased binding of HP1 (genetically identified as Su(Var)2-5) to

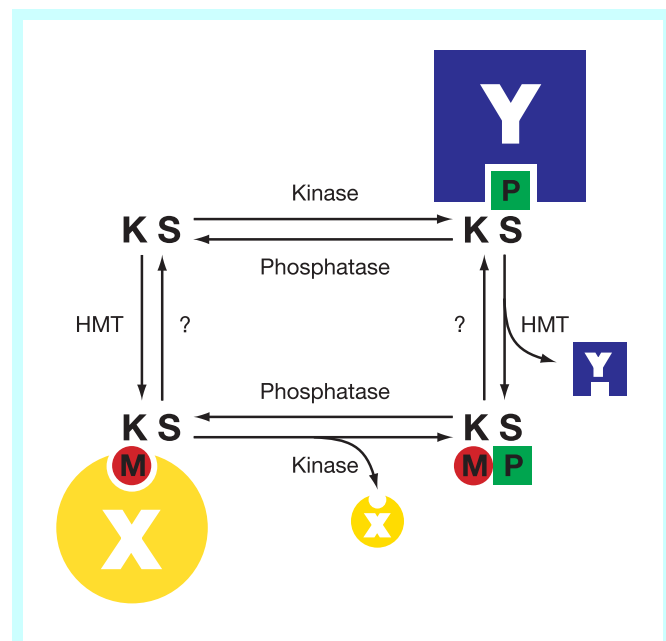


Figure 2 Local binary switches. Post-translational modification of certain amino acids (that is, methylation of lysines [K] or/and phosphorylation of serines [S]) in the core histones results in recruitment of effector proteins (X, Y). In our theoretical model, additional covalent modification of the site adjacent to a previously set covalent mark that has engaged an effector module leads to consecutive loss of binding to that factor. Note that the general concept depicted here for 'methyl/phos switches' could be applicable to other combinations of marks (for example, 'acetyl/phos', 'ubiquitin/phos', etc.). Question marks refer to the yet unknown existence of histone lysine demethylases. HMT, histone methyltransferase (see text for details).

H3 methyl-Lys 9 marks, which are themselves added by a H3 Lys 9 methyltransferase (Su(Var)3-9).

Other histone binary switches

If the general concept of binary 'methyl/phos switches' is correct, we predict that other methyl-mark-specific modules will be identified whose binding might be sensitive to neighbouring phospho-marks. The best-studied methyl marks in the H3 tail (Lys 4, Lys 9 and Lys 27), for example, are all adjacent to novel (Thr 3) (D. F. Hunt and C.D.A., unpublished observations) or previously described (Ser 0 and Ser 28) phospho-acceptors. Whereas the chromodomain of Polycomb preferentially binds the Lys 27 methyl mark²⁴, a module that 'reads' the H3 Lys 4 methyl mark, which has been associated with an 'on' or 'competent' transcriptional state, has yet to be identified. If such a module exists, we predict that its binding to the H3 tail could be regulated by phosphorylation of Thr 3. In addition, an older literature points to Lys 23 in the H3 tail as another methylation site²⁵ (D. F. Hunt and C.D.A., unpublished observations), and phosphorylation of Thr 22 could regulate the biology of this mark. We therefore propose that up to four 'methyl/phos switches' might be operating on the H3 tail alone (Figs 1 and 3).

When considering all four core histones, the number and placement of putative binary switches, inside and outside of histone tails, is provocative (Fig. 3). Lys 79 of H3, for example, represents the first known methylation site in the histone fold regions²⁶. It also lies adjacent to a potential phosphorylation site, Thr 80, and recent genetic screens have implied Lys 79 Thr 80 in a genomic 'silencing cluster'^{27,28}. This cluster involves a corresponding region of H4, Lys 79 and Thr 80, suggesting that 'methyl/phos switches' might regulate the critical interface between H3 and H4 (it is unknown whether Lys 79 of H4 is methylated or/and whether Thr 80 of H4 is phosphorylated). The idea that a 'methyl/phos switch' may operate on a critical interface of the H3:H4 dimer is attractive given the importance of this boundary for nucleosome structure and gene regulation²⁷.

The exercise of identifying all LysSer/Thr or Ser/ThrLys pairs in all major core histones reveals other surprises (Fig. 3). All potential 'switches' are located either in the exposed histone tails, at the edges of helical stretches, or in the connecting loops of the histone fold domains. These sites could therefore all be accessible to post-translational modification and function as potential 'switches'. The well-known ubiquitination sites (Lys 119 and Lys 120 in

human H2A and H2B, respectively) are also adjacent to threonine residues, suggesting that switching mechanisms might influence ubiquitin attachment or removal, and/or the readout of ubiquityl-marks (it is unknown if Thr 120 in H2A or Thr 119 in H2B are phospho-acceptors).

Only three methyl marks, Lys 14 and Lys 36 of H3 and Lys 20 of H4, are not next to known or putative phosphorylation sites (although His 18 in H4 could be a phospho-acceptor regulating H4 Lys 20, see Fig. 1). If the proposed 'methyl/phos switch' hypothesis is correct, it will be interesting to see how these sites and their readout are regulated, and whether their biology differs from methyl marks that are controlled by 'phospho-switching' mechanisms.

Far fewer acetyl-lysine residues than methyl-lysine residues in histones are next to phospho-acceptors (only four out of 14 known acetyl-lysines are neighbouring Ser/Thr residues, but five out of eight methyl-lysines are next to phospho-acceptors). We speculate that this might reflect the lower stability of acetyl-lysine marks, which are readily erased by histone deacetylases (HDAC), in comparison to less labile methyl-lysine marks. In general, this observation might suggest that stable marks might have coevolved with adjacent flexible marks establishing 'local binary switches' for their regulation.

Rules for functional prediction

Our survey of all LysSer/Thr or Ser/ThrLys pairs in the major (human) core histones (Fig. 3), identifies an intriguing pattern that may imply an 'order-first' rule. Putative lysine 'switch' sites that have by now been implicated in binding repressive modules (Lys 9/Ser 10 and Lys 27/Ser 28 in H3 recruiting HP1 and Polycomb, respectively) directly precede the phospho-acceptor. In contrast, the methyl mark comes first in the activating Thr 3/Lys 4 residue pair (discussed above) in the H3 tail. From these observations we hypothesize that 'lysine-first' 'methyl/phos switches' (LysSer/Thr) may dictate a silenced or 'off' state of chromatin. In contrast, 'serine/threonine-first' 'phos/methyl switches' (Ser/ThrLys) may regulate activated or 'on' states of chromatin. The Thr 22/Lys 23 pair in H3 might therefore represent a previously unrecognized activating 'methyl/phos switch'. The situation is, however, less clear at the 'silencing clusters' in H3 (Lys 79/Thr 80) and H4 (Lys 79/Thr 80), and methylation at Lys 79 might represent an exception to the proposed 'order-first' rule. If it is indeed an 'on' mark reducing SIR

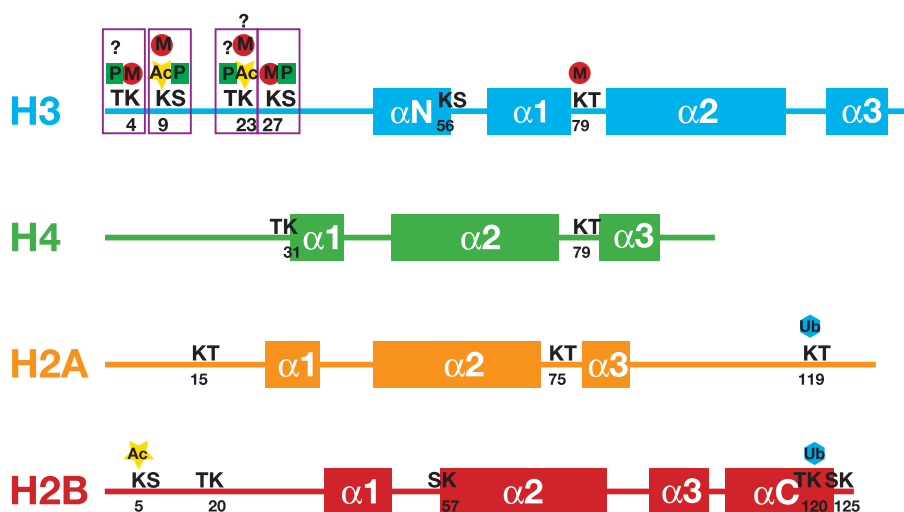


Figure 3 Surveillance of putative 'methyl/phos switches' in core histones. The sequences of all four human core histones were searched for the presence of T/SK and KS/T motifs (T, threonine; S, serine; K, lysine). Known post-translational modifications

of such sites are indicated. Purple boxes frame four likely 'methyl/phos switches' in the H3 tail. Ub, ubiquitination.

repressor binding²⁹, loss of silencing in yeast mutants of H3 Lys 79 and the corresponding HMT, Dot1, must be caused by indirect effects³⁰.

The proposed ‘order-first’ rule might only apply to more-stable marks such as lysine methylation. Acetylation of Lys 9 in H3, for example, is well documented. Several studies have shown that acetyl-Lys 9 synergizes with phospho-Ser 10, and the di-modified, acetyl/phos state of this pair correlates with immediate-early gene activation³¹. However, as indicated earlier, a phospho-switch, proposed in our model to be a ‘release button’ for methyl-binding effector proteins (Fig. 2), may not be needed to ‘eject’ acetyl-binding modules (such as bromodomains), because acetyl-marks are easily erased by HDAC activities. Similar considerations might apply to sites of ubiquitination, as deubiquitinating enzymes exist. Yet we note that Lys 119 of H2A precedes Thr 120, which would be consistent with this ubiquityl mark functioning as an ‘off’ switch. Indeed, recent work shows that ubiquitination of H2A Lys 119 is enriched in the heterochromatic XY sex body during murine spermatogenesis³². In contrast, Lys 120 of H2B is preceded by Thr 119, which would be consistent with this ubiquityl mark functioning as an ‘on’ switch. In support, ubiquitination of H2B governs an activating ‘trans-histone’ regulatory pathway in yeast that involves methylation of Lys 4 and Lys 79 in H3 (ref. 10). Moreover, ubiquitinated H2B is only detected in transcriptionally

active macronuclei of *Tetrahymena* and is absent from transcriptionally silent micronuclei that contain ubiquitinated H2A³³.

How could the proposed ‘order-first’ rule of ‘lysine-first’ (off) versus ‘serine/threonine-first’ (on) switches be implemented? The binding of most modules to their cognate marks is highly asymmetric. For example, the chromodomains of HP1 and Polycomb make contacts to four to six residues N-terminal to the recognized methyl-lysine marks, but ‘see’ only one residue C-terminal to the modification site²⁴. Similarly, the structure of the bromodomain of Gcn5 bound to the H4 Lys 16 acetyl-mark shows major contacts only to the N-terminal face of the binding site. Indeed, asymmetry of binding interfaces is a common phenomenon in cell biology (for example, SH2 domains bind asymmetrically to sequences containing phospho-tyrosine)⁵. We speculate that an asymmetric readout of switch-sites by conserved modules could dictate the biology of these post-translational modifications. Modules that read the ‘on’ switches (like Thr 3 Lys 4^{methyl}) or the ‘off’ switches (like Lys 9^{methyl} Ser 10, Lys 27^{methyl} Ser 28) of the ‘order-first’ rule might therefore recognize opposed and contrary faces of their cognate binding sites.

Modification cassettes

Beyond binary switching mechanisms between two neighbouring or nearby marks, we envision that linear strings of densely clustered modifiable sites in histones might generally not act independently but rather function as discrete information units mediating and relaying different signals. Whereas signature sequence motifs of protein–protein interaction have been defined in many biological pathways, we hypothesize that the combination of dense marks in short clusters, situated at strategic locations in the histones, could form defined ‘modification cassettes’ with distinct biological read-outs depending on their modification state (for example, a ‘cassette’ of 3 marks could have 2^3 possible states if each mark can be modified independently).

Short 'cassettes' where at least three out of five sites separated by no more than one residue can be covalently modified are readily spotted in the H3 and H4 tails (see Fig. 1). The stretch between Lys 9 and Thr 11 of H3, for example, can carry four different covalent marks. Additional clusters of adjacent marks on H3 are centred around Lys 4 and Lys 27. On the H4 tail, 'cassettes' where single modification sites are separated by only one residue can be found. In H4 of most species, Ser 1, Arg 3, and Lys 5 are well-known sites of phosphorylation, methylation and acetylation, respectively. Similarly, Lys 16 and Lys 20, situated more internally in the H4 tail, are sites of acetylation and methylation, respectively, and an older literature suggests that His 18 might be an acid-labile phosphorylation site³⁴.

We reason that the modification of adjacent sites within short clusters of accumulated histone modifications will influence the recognition and binding of modules to their cognate mark (see ‘Local binary switches’ section). Especially since the recognition of short clusters of 5–10 amino acids is a hallmark of the interaction of most modules with their target interaction marks⁵. Furthermore, modules that directly read complex patterns of marks might exist.

Implications beyond histones

Some putative ‘modification cassettes’ of major histones are also found in minor histone variants, possibly arguing for conserved functions. For example, the extreme N terminus of H4 in most species is also found at the very N terminus of most, but not all, H2A variants (Fig. 4a). In contrast, the stretch [AG-GK...], which marks the beginning of the H2A.Z variant in several species, is identical to the extreme N terminus of *Tetrahymena* H4. Could the change of Ser 1 to Ala 1 be linked to the loss of Arg 3 in specific members of both H2A and H4 gene families? Perhaps phosphorylation at Ser 1 in H4 or H2A is linked to methylation at nearby Arg 3 (or vice versa).

p53 shows features of post-translational modification that are

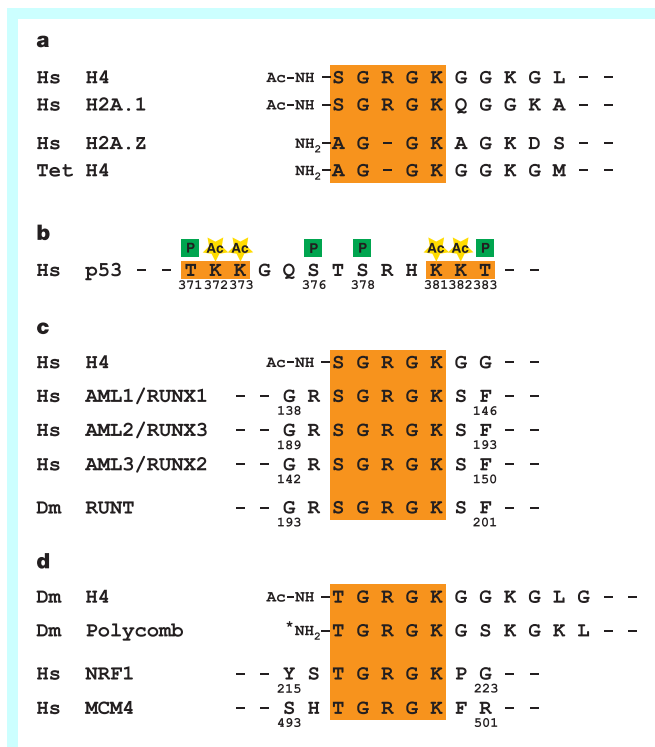


Figure 4 Putative ‘cassettes’ and ‘switches’ in histone and non-histone proteins. **a**, Conservation (and divergence) of a putative ‘modification cassette’ motif in the extreme N terminus of H4 and H2A.1 as well as in H2A.Z, a minor histone variant. Note the change of Ala 1 for Ser 1 concomitant with the loss of Arg 3 exhibited by some H4 and H2A species. **b**, Potential p53 ‘modification cassettes’ and ‘switches’. Post-translational modifications in the C-terminal basic domain of human p53 are denoted. Putative ‘modification cassettes’ are indicated by boxes. **c, d**, Sequence comparison of putative ‘modification cassettes’ motifs in histones and other proteins. The [SGRGK] and [TGRGK] sequence stretches of human and *Drosophila* H4, respectively, were analysed against the Swiss-Prot database (Release 41.11) using the Pattern Search algorithm at http://hits.isb-sib.ch/cgi-bin/hits_patsearch. Note that the indicated cassettes are evolutionarily very conserved in the AML/RUNX protein family as well as in the NRF1 and MCM4 proteins. The asterisk indicates, to our knowledge, that it is not known whether the N terminus of *Drosophila* Polycomb is acetylated.

reminiscent of histones. The basic C-terminal region of this tumour suppressor can be phosphorylated and acetylated on multiple sites (Fig. 4b)^{35,36}, and methylation of p53 in this domain is under active investigation (D. Reinberg, personal communication). Post-translational modifications at eight out of 13 residues not only give rise to a high density of marks but also to a large number of putative combinations of marks. Similar to histones, some cross-talk between different marks on p53 have been described (that is, phospho-marks facilitating consecutive acetylation and acetyl-marks possibly interfering with ubiquitination)^{35,37}. However, whereas discrete functions and signalling events have been associated with some marks (that is, protein stability and DNA binding activity), the biological role of other marks is still debated³⁶. Whether aspects of the biology and function of these post-translational marks on p53 could be explained in the context of 'modification cassettes' and/or by switching mechanisms similar to the ones introduced here for histones remains to be determined.

We suspect that the development of sensitive proteomic approaches will soon lead to the discovery of additional clusters of modifications on other non-histone proteins. Comparison of the sequence motifs of the proposed 'modification cassettes' with other proteins identifies some striking similarities. For example, the extreme N terminus of most H4 proteins [SGRGK...] is found precisely in the AML/RUNX/RUNT family of transcriptional regulators (note the conservation of this motif in sequences from fly to human) (Fig. 4b). Similarly, the *Drosophila* H4 N-terminal motif [TGRGK...] is also present in the NRF and MCM4 proteins of higher organisms and at the very N terminus of *Drosophila* Polycomb (Fig. 4c). While it remains unknown to what extent covalent modifications exist in these short clusters, we speculate that these conserved motifs could represent discrete information units with yet unknown regulatory function. Whether the general concepts presented here for histones might ultimately be applicable to other proteins and signalling cascades remains an intriguing possibility.

Perspective

The ideas presented here provide testable hypotheses of how the extreme high density and versatility of covalent histone modifications could be used for particular signalling purposes. They extend the original histone code hypothesis⁷ and might have relevance beyond histone proteins. Evidence for the occurrence of 'local switches' or 'modification cassettes' will probably come from combined approaches, including biophysical techniques, immunocytological assays, genetic approaches and mass spectrometry. If 'switches' and 'cassettes' indeed operate as discrete functional units, it might be possible to exchange them between separate histones and other proteins.

Whereas the concepts discussed here could explain how the binding and release of effectors is regulated, they do not address the critical issue of whether and to what extent epigenetic inheritance is carried and stored at the level of histone modifications themselves (for example, are certain histone marks part of epigenetic memory systems?). Using H3 as an example, we also recognize that enormous competition is likely to exist between positive and negative effectors over densely marked stretches of an exposed histone domain. Control of such overall 'on-off traffic' of docking modules and the translation of complex modification patterns into meaningful biological responses will probably challenge and surprise biologists for many years to come. □

Received 13 June; accepted 22 August 2003; doi:10.1038/nature02017.

1. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
2. Cheung, P., Allis, C. D. & Sassone-Corsi, P. Signaling to chromatin through histone modifications. *Cell* **103**, 263–271 (2000).
3. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).
4. Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**, 286–298 (2002).

5. Pawson, T., Gish, G. D. & Nash, P. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol.* **11**, 504–511 (2001).
6. Schreiber, S. L. & Bernstein, B. E. Signaling network model of chromatin. *Cell* **111**, 771–778 (2002).
7. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
8. Zhang, K. *et al.* Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by high-accuracy matrix-assisted laser desorption/ionization-time-of-flight, matrix-assisted laser desorption/ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry. *Anal. Biochem.* **306**, 259–269 (2002).
9. Zhang, L., Eugeni, E. E., Parthun, M. R. & Freitas, M. A. Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* **112**, 77–86 (2003).
10. Fischle, W., Wang, Y. & Allis, C. D. Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* **15**, 172–183 (2003).
11. Festenstein, R. *et al.* Modulation of heterochromatin protein 1 dynamics in primary mammalian cells. *Science* **299**, 719–721 (2003).
12. Cheutin, T. *et al.* Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299**, 721–725 (2003).
13. Bannister, A. J., Schneider, R. & Kouzarides, T. Histone methylation: dynamic or static? *Cell* **109**, 801–806 (2002).
14. Ayyanathan, K. *et al.* Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev.* **17**, 1855–1869 (2003).
15. Li, Y., Danzer, J. R., Alvarez, P., Belmont, A. S. & Wallrath, L. L. Effects of tethering HP1 to euchromatic regions of the *Drosophila* genome. *Development* **130**, 1817–1824 (2003).
16. Berger, S. L. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**, 142–148 (2002).
17. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
18. Jacobs, S. A. & Khorasaniadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
19. Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
20. Kellum, R., Raff, J. W. & Alberts, B. M. Heterochromatin protein 1 distribution during development and during the cell cycle in *Drosophila* embryos. *J. Cell Sci.* **108**, 1407–1418 (1995).
21. Minc, E., Allory, Y., Worman, H. J., Courvalin, J. C. & Buendia, B. Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. *Chromosoma* **108**, 220–234 (1999).
22. Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 279–291 (2000).
23. Reuter, G. & Spierer, P. Position effect variegation and chromatin proteins. *Bioessays* **14**, 605–612 (1992).
24. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
25. Waterborg, J. H. Sequence analysis of acetylation and methylation in two histone H3 variants of alfalfa. *J. Biol. Chem.* **265**, 17157–17161 (1990).
26. Lachner, M., O'Sullivan, R. J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116**, 2117–2124 (2003).
27. Park, J. H., Cosgrove, M. S., Youngman, E., Wolberger, C. & Boeke, J. D. A core nucleosome surface crucial for transcriptional silencing. *Nature Genet.* **32**, 273–279 (2002).
28. Thompson, J. S., Snow, M. L., Giles, S., McPherson, L. E. & Grunstein, M. Identification of a functional domain within the essential core of histone H3 that is required for telomeric and HM silencing in *Saccharomyces cerevisiae*. *Genetics* **163**, 447–452 (2003).
29. Ng, H. H., Ciccone, D. N., Morshead, K. B., Oettinger, M. A. & Struhl, K. Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. *Proc. Natl Acad. Sci. USA* **100**, 1820–1825 (2003).
30. van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745–756 (2002).
31. Clayton, A. L., Rose, S., Barratt, M. J. & Mahadevan, L. C. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *EMBO J.* **19**, 3714–3726 (2000).
32. Baarends, W. M. *et al.* Histone ubiquitination and chromatin remodeling in mouse spermatogenesis. *Dev. Biol.* **207**, 322–333 (1999).
33. Davie, J. R., Lin, R. & Allis, C. D. Timing of the appearance of ubiquitinated histones in developing new macronuclei of *Tetrahymena thermophila*. *Biochem. Cell Biol.* **69**, 66–71 (1991).
34. Fujitaki, J. M., Fung, G., Oh, E. Y. & Smith, R. A. Characterization of chemical and enzymatic acid-labile phosphorylation of histone H4 using phosphorus-31 nuclear magnetic resonance. *Biochemistry* **20**, 3658–3664 (1981).
35. Appella, E. & Anderson, C. W. Signaling to p53: breaking the posttranslational modification code. *Pathol. Biol. (Paris)* **48**, 227–245 (2000).
36. Prives, C. & Manley, J. L. Why is p53 acetylated? *Cell* **107**, 815–818 (2001).
37. Brooks, C. L. & Gu, W. Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr. Opin. Cell Biol.* **15**, 164–171 (2003).
38. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).

Acknowledgements We thank all current and past members of the Allis laboratory for their input and support as well as many collaborators with whom we have worked or discussed our ideas. We are especially grateful to T. Jenuwein, D. Reinberg and M. Grunstein for discussions in the process of writing this article. We also thank D. F. Hunt and his laboratory as well as C. M. Barber for their groundbreaking work on the *in vivo* modification patterns of histones that spurred many of the ideas presented here. W.F. is a Robert Black fellow of the Damon Runyon Cancer Research Foundation. Work in the laboratory of C.D.A. is supported by several grants from the NIH.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to C.D.A. (alliscd@mail.rockefeller.edu).

Evolution of complex life cycles in helminth parasites

Geoff A. Parker¹, Jimmy C. Chubb¹, Michael A. Ball² & Guy N. Roberts¹

¹Population and Evolutionary Biology Research Group, School of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, UK

²Division of Applied Mathematics, Department of Mathematical Sciences, University of Liverpool, Liverpool L69 3BX, UK

The fundamental question of how complex life cycles—where there is typically more than one host—evolve in host–parasite systems remains largely unexplored. We suggest that complex cycles in helminths without penetrative infective stages evolve by two essentially different processes, depending on where in the cycle a new host is inserted. In ‘upward incorporation’, a new definitive host, typically higher up a food web and which preys on the original definitive host, is added. Advantages to the parasite are avoidance of mortality due to the predator, greater body size at maturity and higher fecundity. The original host typically becomes an intermediate host, in which reproduction is suppressed. In ‘downward incorporation’, a new intermediate host is added at a lower trophic level; this reduces mortality and facilitates transmission to the original definitive host. These two processes should also apply in helminths with penetrative infective stages, although the mathematical conditions differ.

The explosion of interest in life history theory has generated insights into many strategic aspects of life cycles, such as optimal allocation of resources for growth and reproduction, and the timing of maturity^{1–4}, yet the evolution of parasite life-cycle complexity is fundamental but is seldom considered^{5–7}. Here, we use the term ‘complex life cycle’ to mean movement between two or more hosts in a parasite life history (rather than metamorphosis^{8,9}). Gains from parasitism are emphasized¹⁰, but do not explain transfers between successive hosts. Host switching has risks; nevertheless it is often advantageous^{11,12}.

Helminths show remarkable diversity in mode of life, hosts and life history stages; their co-evolutionary associations with the vertebrates date from the Late Cambrian period to the present day¹³. We define two life-cycle strategies: type 1 lacks asexual reproduction (nematodes, acanthocephalans and most cestodes); type 2 has a larval stage(s) with asexual multiplication (trematodes and some cestodes). Most type 2 cycles in trematodes have infective penetrative stages (miracidia and cercariae). Here we derive a prospective logic for adding new hosts to type 1 cycles; a similar approach can be applied with some modifications to type 2 cycles.

Our ideas centre on spatial overlap of existing and potential new hosts through trophic interactions, although any factors that generate overlap may apply. Host–parasite interactions are often highly specific, but new host–parasite combinations can occur naturally if ecological conditions change. Although survival is usually poor, there are occasional combinations in which the parasite survives and reproduces in the new host¹⁴. These rare events may have profound ecological and evolutionary consequences, and form the basis of our approach to increased life-cycle complexity.

An ancestral, free-living type 1 helminth may have become parasitic by adapting to survive accidental ingestion by a larger organism, and reproducing within its gut. Increasingly complex cycles evolve by the progressive addition of new hosts. Suppose that a direct (one host) cycle has evolved, with the helminth maturing and reproducing in the host; there are three ways that a new host might be added.

Upward incorporation

The first way is where the life cycle extends ‘upwards’, by adding a new host (2) after the existing definitive host (1), which subsequently becomes an intermediate host. Suppose that host 1 is often eaten by a predator 2 of larger mass. Mutants able to survive and reproduce in host 2 avoid mortality, and are favoured if the

costs of the ability to survive in both hosts rather than just one—that is, the costs of generalism—are small. The first such mutants may benefit owing to reduced intraspecific competition, but may face problems finding mates. As parasite fecundity increases with size, increased survivorship generates selection to delay maturity³, allowing further growth and enhanced reproduction in the greater niche space of host 2 (Fig. 1b), an effect that is amplified if helminth size increases with host size^{15,16}. Eventually, reproduction in host 1 may cease if maturity is delayed to a size that can be attained only in host 2 (Fig. 1c). The original definitive host 1 thus becomes an intermediate host, in which helminth size is optimized for transfer to 2 (ref. 17). Upward incorporation generates higher reproductive potential and avoids mortality due to 2.

We envisage a number of steps, outlined in Fig. 1a–c, for a one-host to two-host transition. Evolutionary conditions for upward

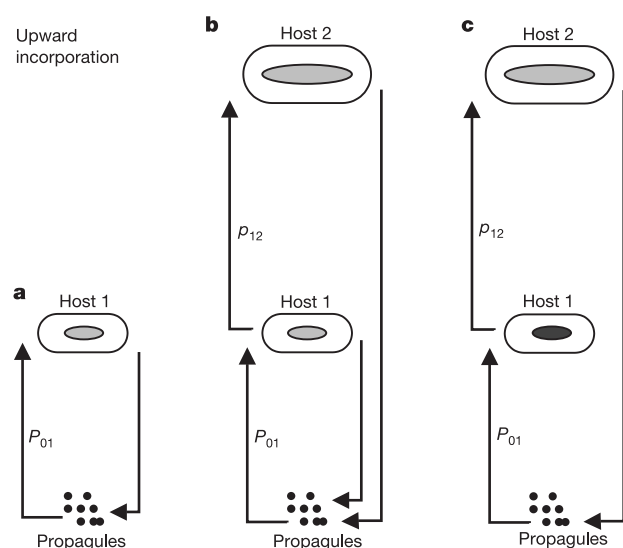


Figure 1 Transition from a one- to a two-host cycle by upward incorporation of a definitive host. **a**, The initial life cycle involves one host; the free-living propagule enters host 1 with probability P_{01} . **b**, This is ingested at rate p_{12} by a predator (potential host 2), resulting (see text) in a ‘flexible’ two-host cycle (the parasite can reproduce in both hosts). **c**, Reproduction in host 1 becomes suppressed, leading to a two-host cycle in which host 1 has become an intermediate host. The grey area indicates adult parasites in definitive hosts; black areas indicate immature parasites in intermediate hosts.

incorporation are derived in the first section of the Methods. We make simplifying assumptions (linear growth, fecundity proportional to size and age-independent mortality); these affect the mathematical conditions, but not the biological conclusions. Below, we write host $1 = i$, $2 = j$.

Parasite adapted to its current host(s)

Step 1 is where the parasite is adapted to mature in its current definitive host i . Let P_{0i} be its probability of survival from its free-living stage 0, possibly through intermediate host(s), to enter host i at a fixed size (for example, size of the free-living stage) at time $t = 0$. P_{0i} depends on survival and transmission to ingestion by i , and on the rate at which i ingests the parasite. If P_{ei} is the probability of establishment in i after ingestion, the probability of successful ingestion by i is $P_{0i}P_{ei} \equiv P_i$. The parasite then grows in i for time t_i^* . Growth stops, and it reproduces at rate B_i until death. B_i increases with body size, and a survival–fecundity trade-off determines t_i^* (ref. 3). Fitness is $w_i = P_i e^{-p_i t_i^*} V_i^*$ (see equation (5)), where mortality of host i occurs at rate p_i , and V_i^* is the parasite's expected total future reproduction at maturity (proportional to the black area in Fig. 2a); that is, the product of birth rate, B_i , and expected adult lifespan, p_i^{-1} . Optimal body size at maturity is $W_i^* = g_i/p_i$; where g_i is the growth rate in i (see equation (9)).

Continued reproduction in an extra higher-trophic-level host

In step 2, host i is a common prey of potential host j . A mutation arises allowing the parasite to survive and reproduce in j , which is a strategy for facultatively exploiting two definitive hosts successively—the ‘flexible cycle’ (Fig. 1b). The phenomenon occurs in present-day helminths with ‘post-cyclic’ hosts¹⁸.

Mortality rate, p_i , has two components: ‘noise’, p_{ni} , resulting in parasite death, and ‘transmission’, p_{ij} , the rate at which host i is consumed by predator j . The mutant has probability P_{ej} of successful establishment in j , where it sustains a mortality rate p_j . If entering j before maturity, it grows at rate g_j , and if it matures, reproduces at rate B_j . We simulate the costs of generalism by reducing the entire fitness of the mutant by factor $(1 - \phi)$, where

ϕ represents the costs of the abilities to survive in both i and j . As i and j are at different trophic levels in a given habitat, ϕ is typically a cost of the ability to survive in a predator at a different trophic level. Initially, reproduction begins at size W_i^* whether in i or j , as the mutant differs only in survivorship.

Assuming that size on entering i is small, the mutation spreads if

$$\frac{p_{ij}}{p_i} \theta > \frac{\phi}{(1 - \phi)} \left(\frac{V_i}{P_{ej} V_{ji}} \right) \quad (1)$$

(see equation (18)), where V_i is the expected lifetime reproductive gains at maturity in i , and V_{ji} the same in j (or gains expected when entering j as an adult); that is, products of reproductive rate, B , and expected lifespan, p^{-1} (Fig. 2). θ depends on mating characteristics: in helminths with separate sexes, only mutants that have matured in i can reproduce in j , and θ is $\theta_s = 1$. For a ‘selfing’ hermaphrodite, θ is θ_h where

$$\theta_h = \left[\frac{e^{(1-\gamma/\mu)} - \gamma/\mu}{1 - \gamma/\mu} \right] \geq 1 \quad (2)$$

and where γ = (growth rate in i /growth rate in j) and μ = (mortality rate of i /mortality rate of j). With separate sexes, if equation (1) is satisfied, as the mutant spreads the condition for increase becomes easier to satisfy, converging asymptotically to equation (2). Equation (2) for selfing hermaphrodites simplifies if γ/μ is small. For small γ/μ values, θ_h approaches $e^1 = 2.72$; for (biologically unrealistic) large γ/μ values, θ_h tends to 1.0.

The inequalities are most likely to be satisfied if ϕ_i , V_i/V_{ji} and γ/μ are small, and P_{ej} and p_{ij}/p_j are large. Thus step 2 is facilitated by: low generalism costs; higher growth and reproductive rates in j than in i ; frequent predation of i by j ; higher mortality of i than j ; easy establishment in j ; and whether single immature parasites ingested by j can grow and reproduce in j (selfing hermaphrodites). Growth rate and reproductive gains in j are likely to be greater than in i (that is, grey area large relative to black area; Fig. 2b), as we expect: longer reproductive lifespan in j than i (j is at a higher trophic level, so $p_i > p_j$); higher reproductive rate in j than i ($B_j > B_i$) because j is likely to be bigger, possibly warm-blooded (and i cold-blooded); and because initially, intraspecific competition in j may be negligible relative to i .

Adult size increases and maturity is delayed

Suppose that the mutation in step 2 has spread to fixation. The next mutations (step 3) change body size at maturity; this increases because of the decreased mortality, higher growth and reproductive prospects in host j (see first section of Methods).

Flexible cycles as in Fig. 1b exist only rarely in helminths. Our explanation is that typically the original definitive host eventually becomes an intermediate host (Fig. 1c). Initially, we expect the probability of reproduction in i gradually to reduce as adult body size increases. Interestingly, several caryophyllaeid and spathebothriid cestodes (*Archigetes*¹⁹, *Cyathocephalus*²⁰) show ‘progenesis’: if transmission to the definitive host is delayed, they mature and reproduce in what is normally the last intermediate host.

Original definitive host becomes an intermediate host

For step 4, adult body size now increases up to the optimum for host j ($W_j^* = g_j/p_j$), but the parasite cannot survive to this size in i (for example, it becomes lethal at size $W_i^m < W_i^*$; see first section of Methods). Reproduction occurs only in j , because the parasite cannot mature in i (Fig. 2c).

Growth arrests before size becomes lethal in host i

At step 5, selection favours growth arrest in i at size W_i^d , just before the parasite reaches lethal size W_i^m (see first section of Methods). It is now adapted to use j exclusively as its definitive host, and i as an intermediate host. The ratio (new adult size in j /old adult size in i) is simply $g_j p_i / g_i p_j = \mu/\gamma$ (compare Fig. 2a, c).

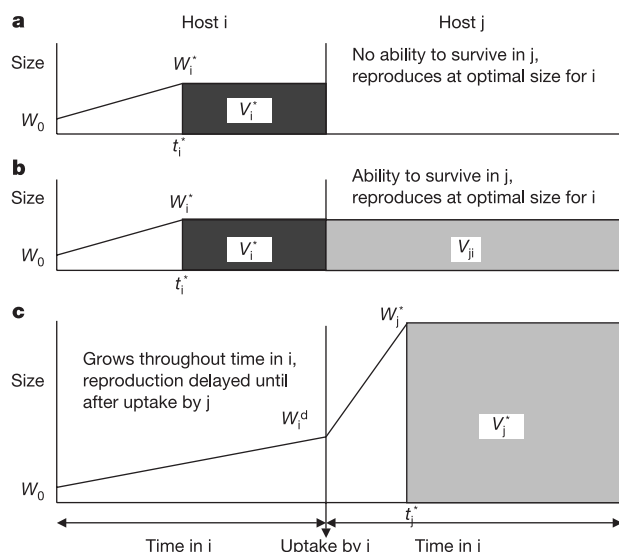


Figure 2 Growth and reproduction in upward incorporation. **a**, One-host cycle: the parasite enters host i at size W_0 and matures after time t_i^* at size W_i^* . Expected reproduction at maturity is proportional to the black area, V_i^* (reproductive rate, B_i , times lifespan, p_i^{-1}). **b**, Flexible two-host cycle: the parasite can also survive and reproduce in host j . Expected reproduction at maturity is proportional to V_i^* plus component V_{ji} in j (proportional to grey area). **c**, Standard two-host cycle: no reproduction occurs in i , and maximum size on entering j is W_i^d . Expected reproduction at maturity (size W_j^*) is V_j^* (proportional to grey area).

There are stability requirements on complex cycles generated by upward incorporation. For example, we assumed that the only way to reach host j is through host i . If transmission from a stage h to j can be direct (for example, j preys on both h and i , or h is free-living and can infect both i and j), costs of generalism can be saved by losing host i . To prevent i being lost, transmission rate p_{hj} must be low: stable linear cycles will have low transmission probabilities between non-successive hosts (h to j), otherwise cycles may reduce in complexity even if they first build up. If generalism costs are low, facultative direct transitions from h to j may occur. The nematode *Pseudoterranova decipiens* may have between three and five hosts in its life cycle, depending on transmission routes²¹.

Downward incorporation

The second way to add a new host is through a downward extension of the life cycle—a new host (2) is added at some stage before the definitive host (1). Suppose that propagules of a helminth normally infect host 1 directly (Fig. 3a). At some point a change occurs: potential host 2, which is lower down the food chain than 1, now also commonly ingests them. If generalism costs are small, and reproduction is impossible in host 2, 2 may become an intermediate or a paratenic host, enhancing transmission to 1. Hence initially, transmission to host 1 may be either direct, or indirect through host 2 (Fig. 3b). Later, the direct pathway may be lost (Fig. 3c). Downward incorporation avoids mortality owing to host 2, and confers benefits of increased transmission to the definitive host 1.

A new intermediate host could evolve between any two successive hosts if it sufficiently increased the parasite's transmission probability between them. The second section of the Methods (assumptions parallel the first section) analyses the simple case where an intermediate host is incorporated into a cycle with one (definitive) host (Fig. 3a–c).

Parasite enters host 1 directly and reproduces in host 1

For step 1, consider a free propagule that does not increase in mass before, with probability P_{01} , it enters only host 1, where it has probability P_{e1} of successful establishment. It grows at rate g_1 until it attains optimal adult size, $W_1^* = g_1/p_1$, then reproduces. Mortality in host 1 is assumed to be noise mortality, at random rate p_1 .

Parasite reaches host 1 directly and indirectly via host 2

Step 2 involves the spread of a mutant that can also infect, establish

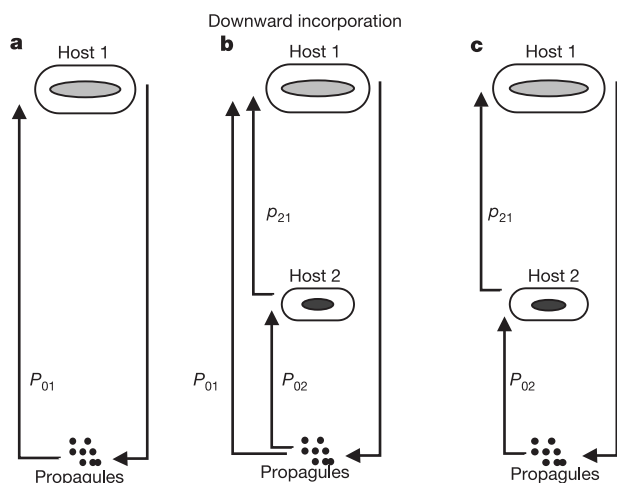


Figure 3 Transition from a one- to a two-host cycle by downward incorporation of an intermediate host. **a**, Initial one-host cycle as in Fig. 1a. **b**, Propagules also enter potential host 2 with probability P_{02} , which is ingested by host 1 at rate p_{21} : infection of host 1 is either direct, or indirect by means of host 2. **c**, Direct transmission to host 1 may later be lost. The grey area indicates adult parasites in definitive hosts; black areas indicate immature parasites in intermediate hosts.

itself, grow at rate g_2 and survive in a smaller host 2 (a common prey item of host 1). Its probability of survival beyond ingestion by host 2 is $P_{02}P_{e2}$, where P_{02} is the probability of surviving to ingestion by host 2, and P_{e2} the probability of establishment in host 2. We assume that if it remains in host 2, it dies before it can attain the optimal size for reproduction in host 1. The mutant again suffers generalism costs, ϕ .

The mutant has two ways to enter the definitive host (Fig. 3b): the direct pathway, to which it is already adapted, and the new indirect pathway, through host 2. Mating occurs only in host 1, where it has no greater difficulty finding mates than a non-mutant. The mutation invades if

$$\left(\frac{p_{21}}{p_2}\right) \left(\frac{P_{02}P_{e2}}{P_{01}}\right) \left[\frac{1 - e^{-(1-\gamma/\mu)p_2 t^m}}{1 - \gamma/\mu}\right] > \frac{\phi}{1 - \phi} \quad (3)$$

(see equation (24)), where p_{21} is the transmission rate (of ingestion of host 2 by host 1), p_{n2} the noise mortality rate in host 2 ($p_2 \equiv p_{21} + p_{n2}$), t^m the maximum time the parasite can survive in host 2, and where $\gamma = g_2/g_1$ and $\mu = p_2/p_1$. Inequality (3) is most easily satisfied when ϕ is small, and p_{21}/p_2 , P_{e2} , P_{02}/P_{01} , $p_2 t^m$ and γ/μ are large. So biologically, step 2 is facilitated by: low generalism costs; frequent predation of host 2 by host 1; easy establishment in host 2; high prospects of infecting host 2 compared with host 1; and longer time to attain lethal size in host 2. Unlike upward incorporation, invasion depends critically on the ratio of P_{02}/P_{01} , there is no dependence on the ratio of birth rates in the two hosts (as all reproduction occurs in host 1), and ratios γ and μ are now likely to be unfavourable for invasion (the reverse applies in upward incorporation).

Growth arrests before the lethal size is attained

With step 3, a mutation spreads that arrests growth in host 2 before the lethal size is achieved.

Parasite no longer enters host 1 directly

The lost ability to exploit the direct pathway (Fig. 3c) cannot be ascribed to avoidance of generalism costs, as survival abilities for both hosts must be retained. A possible reason to lose this ability, at step 4, is that mechanisms of infection and establishment differ for hosts 1 and 2; loss of the direct pathway allows the parasite to specialize on mechanisms for exploiting host 2 early in life, and for 1 later in life, rather than maintaining both simultaneously, which may be energetically more costly.

We now have a standard two-host cycle indistinguishable from that evolved by upward incorporation (compare Figs 1c and 3c). As expected, the fitness terms now correspond exactly with those in step 5 for upward incorporation. Adult size is at an optimum; that is, $W_1^* = g_1/p_1$, so there is no opportunity for mutants inducing small changes to invade. As long as the coefficients obey the same conditions as upward incorporation, size at maturity remains unchanged. The resulting cycle resembles that of many helminths (Fig. 3c).

Lateral incorporation

Many helminths exploit more than one host species at a particular stage in their life cycle. An extra host species may be added at any given point in a cycle, provided that there is niche overlap between the existing and the new host, and generalism costs are sufficiently small. As long as costs are offset by the extra fitness gained, any mutant surviving in the extra host spreads and fixates. Lateral incorporation is a specialist–generalist problem; it does not add to helminth life cycle length and is not considered further here.

Discussion

Our models hinge on the common idea that trophic interactions in food chains lead to the evolution of complex life cycles in helminths with type 1 cycles^{7,13}. The number of hosts per life cycle seldom

exceeds three (that is, one less than the ubiquitous four-level trophic chains supported by theory²² and fact²³). Precise details of trophic interactions will be important: including these could lead to predictions about where and where not to find complex life cycles. Complex cycles necessitate high predation rates on each host, giving high transmission to the next host in the chain. Direct cycles will typically have free-living stages with a high probability of the parasite reaching its definitive host. Ecological stability of food chains generates evolutionary stability of complex life cycles, but should conditions change, shifts to simpler or more complex cycles are entirely possible.

The analysis shows how food-chain evolution, or food-web changes through stochastic processes, fuel helminth life-cycle complexity by increasing the number of hosts, and diverting some of the mortality due to predation into a gain for the parasite. Upward incorporation also boosts reproduction in hosts of greater biomass, relying on the ratios of reproduction in the existing, relative to the incipient, definitive host. Downward incorporation also facilitates transmission-up cycles, relying on the ratios of probability of infection of alternative hosts.

Which process is more likely? Classical views^{13,24} tend towards upward incorporation, but forms of downward incorporation have been advocated strongly¹³. The end result is identical (Figs 1c and 3c): comparative methods²⁵ or molecular biology may be able to answer this question by examining the age of associations with different hosts. Our results show that both routes are possibilities. Growth rates and mortality ratios of the hosts are typically favourable biologically for upward incorporation, but not for downward incorporation; however, downward incorporation is not limited by mode of reproduction (hermaphroditism or separate sexes). Multiple hosts at a trophic level (and cycle stage) can evolve by lateral incorporation. The only obvious selective force acting against ever-increasing life-cycle flexibility relates to costs of generalism, which will range from negligible (physiologically similar hosts) to high (hosts from widely different groups; for example, insect versus mammal).

For downward incorporation, we assume that initially $P_{01} > P_{02}$ (see Fig. 3a, where $P_{02} = 0$). If the helminth had had a higher rate of encounter with host 2 than host 1, 2 is likely to have become the original host. But downward incorporation is most probable if $P_{02} > P_{01}$ (equation (3)). It therefore typically requires that the relative transmission probabilities (P_{01} , P_{02} ; Fig. 3) reverse during evolution of the cycle. Such reversals may occur either because of changes in host evolution (for example, the original host may evolve, thus limiting transmission) or host ecology (habitat is invaded by a potential new host 2 that is preyed on by the original host). Upward incorporation assumes that the food web evolves ahead of the parasite, with new hosts becoming available up the food chain. Downward incorporation assumes that the parasite co-evolves with the original host as it moves up a trophic level, allowing new hosts to become available down the food web.

We have outlined only the basis of increased complexity: subsequent mutations will fine-tune adaptation. For example, optimal growth and maturity may depend on the size and (in flexible cycles) species of host entered (for example, *Bothriocephalus acheilognathi* versus *B. phoxini* in Lake Balaton²⁶). Furthermore, timing of maturity in the definitive host may depend on the size of the larva on entry. These may change conditions for spread and stability. Our analysis relates to fixed transmission rates (or ones that vary about a constant mean). If long-term host population densities and transmission rates change, cycle stability may be challenged and hosts may be lost, as seen in some nematodes²¹. Co-evolution between hosts and parasites may also affect stability. Growth arrest (step 5 of upward incorporation and step 3 of downward incorporation) increases the time available for an intermediate host to be ingested by the next host, by reducing noise mortality either from the host's immune responses, or from lethal levels of parasite growth¹⁷.

Although we have focused on type 1 cycles with trophic transmission, similar logic can be applied to type 2 cycles (having asexual reproduction), and to those with infection by penetrative transmission. Preliminary analyses suggest that the addition of a life history stage with asexual replication need not greatly affect the conditions if fitness remains proportional to the total biomass derived from a given zygote. In penetrative transmission, further modifications become necessary, but conclusions seem qualitatively similar depending on the hypothetical ancestral states. This suggests that our reasoning is relevant to helminths in general, providing that niche overlap occurs between potential hosts.

Helminth life cycles have a more or less fixed sequence of larval stages, most obviously seen in nematodes and acanthocephalans. Cestodes and trematodes show many more variations, but each has a basic theme. This implies that the major incorporation events occurred long ago; however, our analysis raises questions about the dynamics and frequency of host incorporations in the continuing process of helminth adaptive radiation and diversification. □

Methods

Conditions for upward incorporation

Step 1 is reproduction only in host i. Setting $p_i = p_{ni} + p_{ij}$ and $P_i = P_{0i}P_{ei}$, the probability $n_i(t)$ that the given parasite is in i at time t is

$$n_i = P_i e^{-p_i t} \quad (4)$$

Let t_i^* be the time taken to reach maturity after entering i. The expected fitness when reproduction occurs only in i is

$$w_i = \int_{t_i^*}^{\infty} B_i n_i(t) dt = B_i P_i e^{-p_i t_i^*} / p_i \quad (5)$$

that is, the product of probability of surviving to maturity, reproductive (or 'birth') rate, B_i , and expected adult lifespan, p_i^{-1} , in host i.

Growth occurs at rate g_i in i and size at maturity (when growth stops) is W_i . For simplicity we assume that growth is linear from an initial size W_0 ,

$$W_i = g_i t_i + W_0 \quad (6)$$

In a given host, we assume that the birth rate is proportional to parasite size:

$$B_i = b_i W_i \quad (7)$$

For female parasites, this is reasonable. The optimal age of maturity has

$$dw_i/dt_i^* = (P_i b_i / p_i) [g_i - p_i (g_i t_i^* + W_0)] e^{-p_i t_i^*} = 0; \text{ hence } t_i^* = 1/p_i - W_0/g_i \quad (8)$$

Substituting equation (8) into (6), the optimal size at maturity is

$$W_i^* = g_i / p_i \quad (9)$$

The next step (step 2) is a flexible cycle allowing survival and reproduction in host j. We similarly define $n_j(t)$. We assume that $p_j < p_i$ so that at time t

$$n_j = P_i P_{ej} [p_{ij} / (p_i - p_j)] [e^{-p_i t} - e^{-p_j t}] \quad (10)$$

The mutant able to survive and continue development and reproduction in potential host j has a birth rate in j of $B_j = b_j W_j$. Mutant fitness is

$$w_{ij} = (1 - \phi)(w_i + w_{ja} + w_{jb}) \quad (11)$$

where w_{ja} is the expected reproduction of a mutant ingested by j before t_i^* , and w_{jb} is the expected reproduction of a mutant ingested after t_i^* .

For the calculation of w_{ja} , suppose that host j eats host i in the interval $[t_{ij}, t_{ij} + \delta t]$ where $t_{ij} < t_i^*$. The probability that a given mutant is able to establish in j in this interval is $P_{ej} p_{ij} n_i(t_{ij}) \delta t$. Its size is $g_i t_{ij} + W_0$. It takes extra time t' to reach maturity size W_j :

$$t' = (W_j - W_0) / g_j - g_i t_{ij} / g_j \quad (12)$$

Thus

$$w_{ja} = \int_0^{t_i^*} B_j P_{ej} n_i(t) e^{-p_j t'} dt / p_j = [B_j P_{ej} p_{ij} / p_j] \int_0^{t_i^*} dt (e^{-p_i t} - e^{-p_j t}) \quad (13)$$

where the integrating variable t is t_{ij} . Setting $Z = P_i P_{ej} p_{ij} / p_j$,

$$w_{ja} = B_j Z (e^{-p_i (W_i - W_0) / g_i} - e^{-p_i t_i^* + p_j (W_i - W_0) / g_j}) / (p_i - p_j g_i / g_j) \quad (14)$$

For the calculation of w_{jb} , the number of mature parasites that have entered j at time t is $n_j(t)$ minus those that would have entered j before maturing and which have not died; that is, $n_j(t_i^*) e^{-p_j (t - t_i^*)}$. This is $n_j'(t) = n_j(t) - P_i P_{ej} p_{ij} (e^{-p_i t} - e^{-p_i t_i^*}) e^{-p_j (t - t_i^*)} / (p_i - p_j)$. Thus w_{jb} is

$$\int_{t_i^*}^{\infty} dt B_k n_j'(t) = B_k Z p_j [e^{-p_i t_i^*} / p_j - e^{-p_i t_i^*} / p_i] / (p_i - p_j) = B_k Z [e^{-p_i t_i^*} / p_i] \quad (15)$$

Compare this with equation (5), where B_k is the birth rate in j of a parasite that has matured in i.

The mutant has the same size at maturity in j and i; that is, $g_i / p_i = W_j^* = W_i^* \equiv W^*$, so we expect $B_k = B_j$. However, $g_i \neq g_j$, as the parasite's growth differs in i and j

$$w_{ja} + w_{jb} = B_j Z e^{-p_i t_i^*} [e^{(1 - \gamma/\mu) p_i t_i^*} - \gamma/\mu] / (p_i [1 - \gamma/\mu]) \quad (16)$$

where $\gamma = g_i/g_j$ and $\mu = p_i/p_j$. This we can write as $B_j Z e^{-p_i t_i^*} f(\gamma/\mu, p_i t_i^*)$, where $f(\gamma/\mu, z) = (e^{(1-\gamma/\mu)z} - \gamma/\mu)/(1 - \gamma/\mu)$. $f(\gamma/\mu, z)$ is a monotonically decreasing function of γ for $\gamma \geq 0, z > 0$. We can neglect w_{ji} in comparison with w_{ja} when $p_i \gg p_j$ and $g_i < g_j$.

Regarding conditions for spread, for parasites that must mate to reproduce we ignore w_{ja} , relating to mutants arriving in j before maturity. So for spread, we require $(1 - \phi)(w_i + w_{ji}) > w_{ji}$, giving

$$P_{ej} p_{ij} / p_i > \phi / (1 - \phi) \beta \quad (17)$$

where $\beta = b_i/b_j$. If equation (17) is satisfied the mutation spreads and w_{ja} increases from zero as mating prospects in j increase, accelerating the rate of spread.

For hermaphrodite helminths capable of selfing, both w_{ja} and w_{ji} must be considered. From equation (16), $(P_{ej} p_{ij} / p_i)(e^{(1-\gamma/\mu)p_i t_i^*} - \gamma/\mu)/(1 - \gamma/\mu) > \phi/(1 - \phi)/\beta$. If the initial size $W_0 \ll W_j$, $t_i^* \approx 1/p_i$, $(P_{ej} p_{ij} / p_i)f(\gamma/\mu, 1) > \phi/(1 - \phi)/\beta$. The maximum value of $f(\gamma/\mu, 1)$ is when $\gamma/\mu = 0$, when $f(0, 1) = e$. So for spread

$$(P_{ej} p_{ij} / p_i) > \phi / (1 - \phi) \theta \beta \quad (18)$$

where $1/e < \theta < 1$, equivalent to equation (1), as $V_i/V_{ji} = (b_i/p_i)/(b_j/p_j)$.

Step 3 is where optimal adult size increases and maturity is delayed. If the mutation spreads to fixation, size at maturity, W^* , (considered as a variable) increases if the derivative of $w_i + w_{ja} + w_{ji}$ (see equation (11)) with respect to W^* at $W^* = g_i/p_i$ is positive. t^* is a function of W^* .

At step 1, w_i is optimal at W_i^* . The differential coefficients of w_i and w_{ji} at W_i^* therefore equal 0. Writing $w_{ja} \approx Z B_j [-e^{-p_i t^*} + e^{-g_i p_i t^* / g_i}] / (p_i - p_i g_i / g_j)$,

$$\partial w_{ja} / \partial W_i^* = Z B_j \{ -e^{-p_i t^*} (-1 + W^* p_i / g_i) + e^{-g_i p_i t^* / g_i} (1 - W^* p_i / g_i) \} / (p_i - p_i g_i / g_j) \quad (19)$$

The first part of equation (19) is zero at step 2 but becomes positive as W^* increases and if (as expected) $p_i/p_j = 1/\mu < 1$, $g_i/g_j = 1/\gamma > 1$, then $\partial w_{ja} / \partial W_i^*$ is positive. If the flexible cycle (step 2) has evolved, adult size, W^* , increases.

At step 4 the host i becomes an intermediate host. We propose that maturity size, W^* , keeps increasing for $W^* < g_i/p_i$, readily sustainable by host j , but that the parasite dies if it reaches a size W_i^m (where $g_i/p_i > g_i/p_i = W_i^*$) in i .

Does W^* increase for $W^* < W_i^m$? From equation (18), $\partial w_{ja} / \partial W^*$ is positive. Usually $w_{ja} \gg w_{ji}$ and $w_{ji} > w_i$ if β is large. If so, we only need to consider w_{ja} . Does W^* increase for $W_i^m \leq W^* < g_i/p_i$? We model this by taking $p_{ni}(t)$ as constant until $g_i t = W_i^m - W_0$. Then $n_i(t) = P_i e^{-p_i t}$ for $t < (W_i^m - W_0)/g_i$ (but $n_i(t) = 0$ otherwise). Adult size, W^* , is greater than W_i^m and the only non-zero fitness term is w_{ja} .

Suppose j eats i at time t_{ij} . The probability that the parasite establishes in j is $P_{ej} p_{ij} n_i(t_{ij})$. The probability of surviving to reach W^* is $P_{ej} p_{ij} n_i(t_{ij}) e^{-p_i t^*}$, where $g_i t^* + g_i t_{ij} + W_0 = W^*$. We write $t^m = (W_i^m - W_0)/g_i$; that is, the time taken to grow to W_i^m in host i . Then

$$w_{ja} = (B_j / p_j) \int_0^{t^m} dt_{ij} P_{ej} p_{ij} n_i(t_{ij}) e^{-p_i [W^* - W_0 - g_i t_{ij}] / g_i} \\ = B_j Z e^{-p_i t^*} [(1 - e^{-(p_i - g_i p_i / g_j) t^m}) / (p_i - g_i p_i / g_j)] \quad (20)$$

This has a maximum at $W^* = W_j^* = g_j/p_j$, corresponding to step 1 (equation (9)); the ratio of the adult sizes (step 4/step 1) is $g_j p_i / g_i p_j = \mu/\gamma$. For $W^* < g_j/p_j$, w_{ja} increases monotonically with W^* , so that W^* increases until it reaches its maximum. So now j is the definitive host and i the intermediate host.

The next step is step 5, where growth arrests at a size just less than W_i^m in host i . In step 4 some parasites die because they reach the lethal size W_i^m in i . It obviously pays to stop growth in i at a size $W_i^d < W_i^m$. Suppose that $dW/dt_i = g_i$ for $W_i < W_i^d$ (but $dW/dt_i = 0$ otherwise). In this model, we expect W_i^d to be as large as possible; that is, just smaller than W_i^m . Call the time taken to reach the size for growth arrest or dormancy $t^d = (W_i^d - W_0)/g_i$. Let $w_{ja} = w'_{ja} + w''_{ja}$, where w'_{ja} is the total number of eggs produced if i is ingested before t^d , and w''_{ja} is the eggs produced if ingested after t^d . From equation (20): $w'_{ja} = B_j Z e^{-p_i t^*} [(1 - e^{-(p_i - g_i p_i / g_j) t^d}) / (p_i - g_i p_i / g_j)]$. If i is eaten by j at time $t_{ij} > t^d$, it takes an extra time to reach maturity = $(W^* - W_i^d)/g_j$. Thus $w''_{ja} = (p_{ij} P_{ej} B_j / p_i p_j) e^{-p_i t^*} e^{-p_i t^d (1 - \gamma/\mu)}$. This has the same dependence on W^* as w'_{ja} , so the maximum value of W^* is still g_j/p_j .

Steps 1–5 give a transformation of i into an intermediate and j into a new definitive host. As i and j are unspecified, evolution of further hosts by upward incorporation can be envisaged, starting again at step 1. The analysis applies to incorporation of a single host j into a simple free-living cycle, or to the addition of a new definitive host j when there are a number of prior hosts ...h, i, in the cycle.

Conditions for downwards incorporation

The notation and logic follow the previous section.

Step 1 is where a parasite enters host 1 directly and reproduces only in host 1. At time 0 the parasite enters host 1. The probability that it is alive in host 1 at time t is

$$n_1(t) = P_{01} P_{e1} e^{-p_1 t} \quad (21)$$

and its expected fitness is

$$w_1 = B_1 P_1 e^{-p_1 t_i^*} / p_1 \quad (22)$$

where t_i^* is the optimal time to switch from growth to reproduction, and B_1 is the egg production rate of the adult parasite. The parasite grows at rate g_1 from initial size W_0 , and at maturity has size $W_1^* = g_1 t_i^* + W_0$. B_1 is proportional to adult size: $B_1 = b_1 W_1$. Hence $t_i^* = 1/p_1 - W_0/g_1$ and $W_1^* = g_1/p_1$.

For step 2, the parasite reaches host 2 both directly, and indirectly by means of host 2. A mutant that can survive in host 2 has size W_1^* at maturity, as before. We define a size W_2^m at which (we assume initially) the mutant dies in host 2. $W_2^m < W_1^*$, so that it cannot reproduce in host 2.

Calling time $t = 0$ the instant of ingestion by host 2 and $t^m = (W_2^m - W_0)/g_2$ the

maximum survival time in host 2 (that is to attain size W_2^m), the probability that the mutant is alive and in host 2 at time t is $n_2(t) = P_{02} P_{e2} e^{-(P_{21} + p_{22})t}$ for $t^m < t^d$, $n_2(t) = 0$ for $t^m < t^d$.

Some mutants are ingested directly by host 1, whereas others first enter host 2, which may then be ingested by host 1 where they reproduce. Thus mutant fitness is $w = (1 - \phi)(w_1 + w'_1)$, where w'_1 is the component for parasites passing through host 2. To calculate w'_1 , suppose host 2 is eaten by host 1 in the interval $[t, t + \delta t]$ where $t < t^m$. The probability of surviving in host 1 is $p_{21} P_{e1} n_2(t)$, where $n_2(t)$ is the probability of being alive in host 2 at time t , at which time the parasite has size $W_0/g_2 t$. To reach maturity will take time $(W_1^* - W_0 - g_2 t)/g_1 = t_1^* - g_2 t/g_1$. The probability of the parasite surviving during this time is $e^{-p_1 (g_1 t_1^* - g_2 t)/g_1}$. Its expected lifetime number of eggs if it survives this long is B_1/p_1 . Thus

$$w'_1 = [B_1/p_1] [P_{02} P_{e2} P_{e1} P_{21} \int_0^{t^m} e^{-p_1 t - p_1 (g_1 t_1^* - g_2 t)/g_1} dt] \\ = B_1 P_{02} P_{e1} P_{e2} [g_1 e^{-p_1 t_1^*} p_{21} / (g_1 p_2 - g_2 p_1)] [1 - e^{-(p_2 - g_2 p_1/g_1) t^m}] \quad (23)$$

This mutation invades if $w'_1/w_1 > \phi/(1 - \phi)$; that is, if compare with upwards incorporation:

$$\{P_{02} P_{e2} p_{21} / P_{01}\} [1 - e^{-(1-\gamma/\mu) p_2 t^m}] / p_2 (1 - \gamma/\mu) > \phi / (1 - \phi) \quad (24)$$

where $\gamma = g_2/g_1$ and $\mu = p_2/p_1$ (see equation (3)).

Step 3 is growth arrest. Here, a mutant arises that arrests growth in host 2 at size $W_2^d < W_2^m$: as long as W_2^d is only slightly smaller than W_2^m , this stage is obviously advantageous over stage 2 because parasites do not die when they reach size W_2^m .

Step 4 is loss of the direct pathway. Here, the parasite may lose the direct pathway so that fitness component w_1 is lost. Fitness is now equal to step 5 of the first section of the Methods, putting $i = 2$, $j = 1$, where $w'_{ja} = w'_1$ and w''_{ja} is the additional fitness owing to arrested growth in host 2.

Received 10 July; accepted 19 August 2003; doi:10.1038/nature02012.

- Roff, D. A. *The Evolution of Life Histories: Theory and Analysis* (Chapman & Hall, New York, 1992).
- Kozłowski, J. Optimal allocation of resources to growth and reproduction: implications for age and size at maturity. *Trends Ecol. Evol.* **7**, 15–19 (1992).
- Charnov, E. L. *Life History Invariants* (Oxford Univ. Press, Oxford, 1993).
- Gemmell, A. W., Skorping, A. & Read, A. F. Optimal timing of first reproduction in parasitic nematodes. *J. Evol. Biol.* **12**, 1148–1156 (1999).
- Poulin, R. *Evolutionary Ecology of Parasites* (Chapman & Hall, London, 1998).
- Poulin, R. & Cribb, T. H. Trematode life cycles: short is sweet? *Trends Parasitol.* **18**, 176–183 (2002).
- Brown, S. P., Renaud, F., Guegan, J. F. & Thomas, F. Evolution of trophic transmission in parasites: the need to reach a mating place? *J. Evol. Biol.* **14**, 815–820 (2001).
- Redi, F. *Experimenta circa generationem insectorum, ad nobilissimum vivum Carolum Dati* (Amstelodami, 1671).
- Eberman, B. Evolution in organisms that change their niches during the life-cycle. *Am. Nat.* **139**, 990–1021 (1992).
- Price, P. W. *Evolutionary Biology of Parasites* (Princeton Univ. Press, Princeton, 1980).
- Esch, G. W. & Fernández, J. C. *A Functional Biology of Parasitism* (Chapman & Hall, London, 1993).
- Morand, S., Robert, F. & Connors, V. A. Complexity in parasite life cycles: population biology of cestodes in fish. *J. Anim. Ecol.* **64**, 256–264 (1995).
- Baer, J. G. *Ecology of Animal Parasites* (Univ. Illinois Press, Urbana, 1952).
- Moravec, F. *Parasitic Nematodes of Freshwater Fish in Europe* (Academic/Kluwer, Prague/Dordrecht, 1994).
- Harvey, P. H. & Keymer, A. E. Comparing life histories using phylogenies. *Phil. Trans. R. Soc. Lond. B* **332**, 31–39 (1991).
- Morand, S. Life-history traits in parasitic nematodes: a comparative approach for the search of invariants. *Funct. Ecol.* **10**, 210–218 (1996).
- Parker, G. A., Chubb, J. C., Roberts, G. N., Michaud, M. & Milinski, M. Optimal growth strategies of larval helminths in their intermediate hosts. *J. Evol. Biol.* **16**, 47–54 (2003).
- Nickol, B. B. in *Biology of the Acanthocephala* (eds Crompton, S. W. T. & Nickol, B. B.) (Cambridge Univ. Press, Cambridge, 1985).
- Mackiewicz, J. S. Caryophyllidea (Cestoidea): a review. *Exp. Parasitol.* **31**, 417–512 (1972).
- Wisniewski, L. *Cyathocephalus truncatus* Pallas, II. Allgemeine Morphologie. *Bull. Int. Acad. Polon. Sci. Lett. Cracovie. Class Sci. Math. Nat.* **B 3**, 311–327 (1933).
- Anderson, R. C. *Nematode Parasites of Vertebrates* 2nd edn (CABI, Wallingford, 2000).
- Matsuno, K. & Ohno, N. How many trophic levels are there? *J. Theor. Biol.* **180**, 105–109 (1996).
- Briand, F. & Cohen, J. E. Environmental correlates of food chain length. *Science* **238**, 956–960 (1987).
- Stunkard, H. W. Interrelationships and taxonomy of the digenetic trematodes. *Biol. Rev.* **21**, 148–158 (1946).
- Brooks, D. R. & McLennan, D. A. *Parascript: Parasites & the Language of Evolution* (Smithsonian Institution, Washington, 1993).
- Molnár, K. *Bothriocephalus phoxini* sp. N. (Cestoda, Pseudophyllidea) from *Phoxinus phoxinus* L. *Folia Parasitol.* **14**, 83–86 (1968).

Acknowledgements We thank the NERC for funding. Pilot versions of some of these ideas were first analysed by G.A.P. while supported by a Nuffield Research Fellowship in 1983.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.A.P. (gap@liverpool.ac.uk).

Programmable computing with a single magnetoresistive element

A. Ney, C. Pampuch, R. Koch & K. H. Ploog

Paul-Drude-Institut für Festkörperelektronik, Hausvogteiplatz 5-7,
D-10117 Berlin, Germany

The development of transistor-based integrated circuits for modern computing is a story of great success. However, the proved concept for enhancing computational power by continuous miniaturization is approaching its fundamental limits. Alternative approaches consider logic elements that are reconfigurable at run-time to overcome the rigid architecture of the present hardware systems¹. Implementation of parallel algorithms on such 'chameleon' processors has the potential to yield a dramatic increase of computational speed, competitive with that of supercomputers². Owing to their functional flexibility, 'chameleon' processors can be readily optimized with respect to any computer application. In conventional microprocessors, information must be transferred to a memory to prevent it from getting lost, because electrically processed information is volatile. Therefore the computational performance can be improved if the logic gate is additionally capable of storing the output. Here we describe a simple hardware concept for a programmable logic element that is based on a single magnetic random access memory (MRAM^{3,4}) cell. It combines the inherent advantage of a non-volatile output with flexible functionality which can be selected at run-time to operate as an AND, OR, NAND or NOR gate.

MRAMs consist of giant-magnetoresistive^{5,6} (GMR) or tunneling magnetoresistive⁷ (TMR) elements which are composed of two magnetic layers (1 and 2 in Fig. 1). These are separated by a non-magnetic spacer or a tunnelling barrier that suppresses magnetic coupling, and guarantees that the magnetizations M_1 and M_2 of the two layers can be rotated independently. The resistance of a magnetoresistive (MR) element depends on the relative orientation of M_1 and M_2 and is considerably lower for parallel alignment. The two magnetic states—parallel and antiparallel, characterized by low and high resistance—can be identified with a logical 1 and 0, respectively. For writing a bit, a current with a magnetic field that is sufficient to rotate the magnetization of one of the two layers is required. Since the direction of the magnetization is maintained when the current is turned off, the information is non-volatile and can be read out repeatedly by measuring the resistance of the MR element without the periodic refreshing needed in CMOS (complementary metal oxide semiconductor) technology.

Furthermore, MR elements offer enhanced logic abilities compared to the rigid architecture of transistor-based logic elements where the functionality is fixed by the wiring. In two recent studies it was proposed to form a programmable spin-logic element by altering the switching threshold of an MR element⁸, or by combination of several MR elements^{8,9}. However, we will show that a single MR element with two independent input lines (A and B in Fig. 1) is already sufficient to provide three different functionalities: the conventional storage bit as well as AND and OR gates selectable simply by the addressing ('set') procedure. By adding a third independent input line C the logic gates NAND and NOR can also be realized.

The MR element required for the proposed logic gate (Fig. 1) consists of two magnetic layers with different coercive fields H_{c1} and H_{c2} where $H_{c1} < H_{c2}$. The input lines A and B are operated with positive or negative currents $\pm I_A$ and $\pm I_B$ of equal magnitude. Each of the two currents alone is not able to generate a magnetic field sufficient to reverse M_1 , because the coercive fields H_{c1} and H_{c2} are

too large. However, I_A and I_B together are able to rotate M_1 but not M_2 . For rotation of M_2 an additional current $\pm I_C$ on input line C is necessary. We now discuss the realization of the AND and OR functionalities by particularly using the non-volatility of MR elements (see also Fig. 2).

AND gate. See blue arrows in Fig. 2a. We assume that the magnetization M_2 of the lower layer points to the right; according to the chosen convention it is positive ($+M_2$). Antiparallel (parallel) alignment of the upper layer, that is $-M_1$ ($+M_1$), corresponds to a logical 0 (1) as output of the MR element. Magnetization with negative (positive) sign is accomplished by negative (positive) currents at the input lines ($\pm I_A$, $\pm I_B$) identifiable with logical 0 (1) of the inputs. Before the logic operation the system is set to the antiparallel configuration, that is $-M_1/+M_2$, by applying 0 at both inputs A and B. This configuration corresponds to an output 0. In the next step, the logic operation takes place: inputs A and B are addressed independently with a 0 or a 1. The direction of the magnetization remains unchanged if 0 is applied at both inputs A and B. The same holds if 1 is applied either at input A or at input B. The magnetization of the upper layer can only be switched to $+M_1$ by applying a logical 1 at both inputs A and B. Only then does the output change to 1 (parallel configuration). The corresponding logic table given in Fig. 2a (blue entries) can easily be recognized as the binary logic AND function.

OR gate. See red arrows in Fig. 2a. To realize the logic functionality of an OR gate the MR element has to be set to a different, but again well-defined, initial state before the logic operation: the magnetization of the upper layer is set to $+M_1$ by applying a 1 simultaneously to inputs A and B; the magnetization of the lower layer remains unchanged at $+M_2$. The parallel configuration $+M_1/+M_2$ with the output 1 is maintained even when the input currents are turned off—a direct consequence of the non-volatility of MR elements. The subsequent logic operation proceeds as follows: applying a 1 to both inputs A and B repeats the set sequence and the output remains 1. With only one input being 1 and the other 0, the resulting field is not sufficient to rotate M_1 , thus keeping the output at 1. Only after applying negative currents (that is 0) to both inputs A and B, $+M_1$ changes to $-M_1$. Then the two magnetizations are aligned antiparallel, which corresponds to an output 0. The resulting logic table is also given in Fig. 2a (red entries) and represents the logic OR function.

The preceding discussion illustrates the inherent properties of MR elements. If they are employed as spin-logic gates, the operation has to proceed in two steps, namely the 'set operation' followed by the 'logic operation'. The first step (where both inputs are addressed together) allows us to choose between the AND and OR functionality by setting the magnetization of the upper layer to $-M_1$ or $+M_1$, respectively. This means that for the different functionalities AND

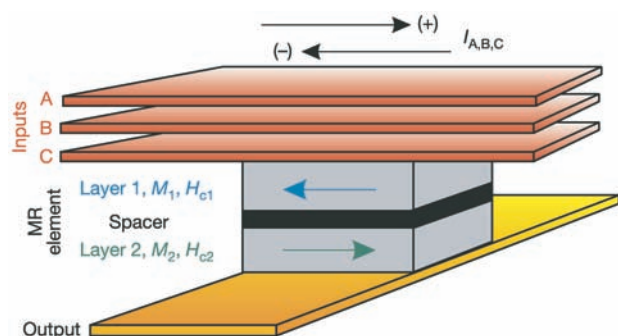


Figure 1 Schematics of a programmable spin-logic device based upon a single MR element with three independent input lines A, B and C and an output line.

and OR the output has to be preset to 0 and 1, respectively, before the logic operation. Then in a second step both inputs are addressed separately and the preselected logic operation is performed.

Although at first sight a two-step procedure appears to be a drawback in speed compared to CMOS technology, our concept offers several advantages: (1) since the output is non-volatile, it can be repeatedly read out via the GMR or TMR effect (analogous to an MRAM). Consequently, a time-consuming transfer of the output to a cache memory is not required. (2) With a non-volatile logic gate the different stages of a computation need not necessarily be synchronized to a single clock. Therefore the gates can be operated asynchronously, which is expected to yield low power circuits¹⁰ and support parallel computation. (3) The set procedure makes run-time programmability feasible, which is expected to increase the computation efficiency by up to three orders of magnitude owing to massive parallel and pipelined execution². The proposed logic element therefore represents the simplest solution of a programmable spin-logic gate based on a single conventional MRAM cell. By using only the 'set' and 'read' steps, the device acts as an ordinary MRAM bit.

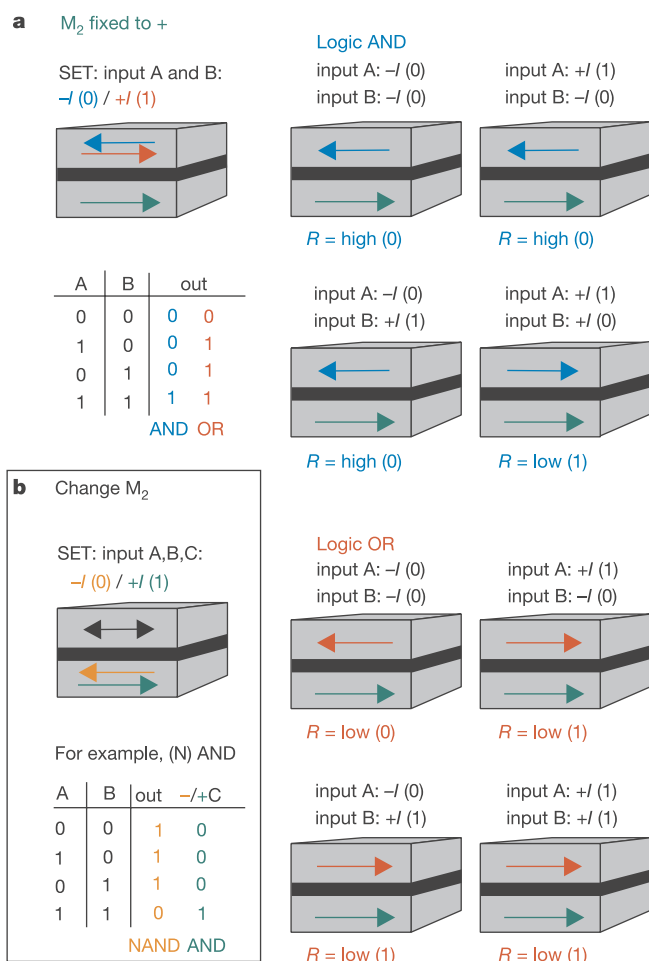


Figure 2 Principle of operation of the MR-element employed as logic gate. **a**, Set operation by switching the magnetization of the upper layer (M_1) to preselect between the logic functions AND (blue arrow) and OR (red arrow) by simultaneously activating inputs A and B with currents $+I$ or $-I$. The subsequent logic operation for AND (blue) and OR (red) is executed by the input currents $\pm I$ of lines A and B leading to different magnetization alignments with resulting high or low resistance (R) identified with the output. The logic input and output values are summarized in the logic table. **b**, Negation of the output (orange) is achieved by reversing the magnetization of the lower layer (M_2) by simultaneously addressing all three input lines. The subsequent logic operation proceeds analogously to **a**.

NAND and NOR gates. See Fig. 2b. So far we have assumed that the lower layer 2 has the magnetization $+M_2$. The logic functionality (AND, OR) of the MR element is solely defined by the relative orientation between the magnetization of the upper layer M_1 and the input currents I_A and I_B . Reversal of M_2 only affects the value of the output changing the magnetoresistance from low to high, or vice versa. Reversal of M_2 therefore corresponds to a negation and turns AND to NAND and OR to NOR (see Fig. 2b). For rotation of M_2 currents of equal sign have to be applied to all three input lines to overcome the coercive field H_{c2} .

We illustrate the complete sequence of a logic AND operation followed by a NOR operation in Fig. 3: in a first 'set' step $-M_2$ or $+M_2$ is selected by activating all three input lines—(i) in Fig. 3—leading to a negation of the output or not. In a second set step (ii), layer 1 is programmed by simultaneously addressing the input lines A and B, yielding either the (N)AND or (N)OR functionality; input line C is not addressed. The device is now pre-programmed. After the set-procedure the 'operation' step takes place by independently using lines A and B as logic inputs. The last step is to read the output by measuring the magnetoresistance between a single input line, for example, C, and the output line (see Fig. 1). The resulting signal can be amplified if required and distributed to other MR elements, for example, by repeated read-out.

The new concept for a universal, non-volatile spin-logic gate, presented here, is based on a single MR element with three input lines and one output line. Whereas an MR element possesses four distinct magnetization states, the output distinguishes only between two states, namely antiparallel and parallel, irrespective of the individual magnetization directions. Each of the four states, however, represents a different logic functionality, making the MR element suitable to serve either as AND, OR, NAND, or NOR gate or as a conventional MRAM cell. The logic functionality has to be predefined before the logic operation is executed.

The MR element has several advantages compared to common logic devices: (1) a single MR element is sufficient to realize and store the four basic logic functionalities. Since at least two transistors are needed for each CMOS logic gate, the integration density is effectively increased. (2) The output is non-volatile and repeatedly readable without refreshing, which reduces the heat evolution. (3) After the logic operation the device is again in a defined state for a logic operation which can be directly (that is, without a further set-

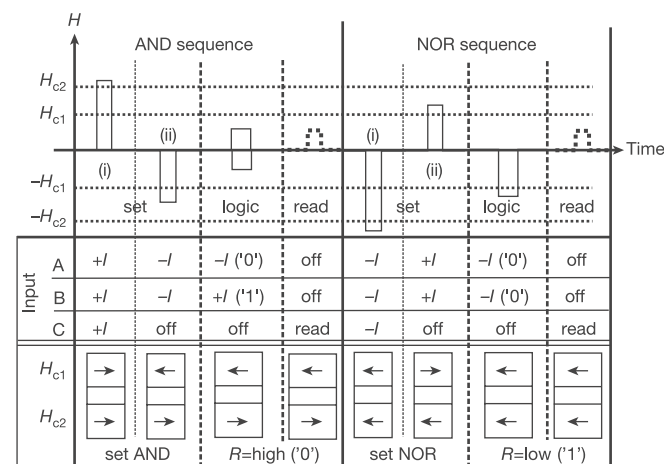


Figure 3 Sequence of set, logic, and read operations for AND and NOR. Two initial set steps are performed: (i) to enable/disable negation of the output by addressing all three inputs, and (ii) to choose between (N)AND and (N)OR using only two inputs. The third step is the logical operation analogous to Fig. 2. The last step is the read-out of the output via the GMR/TMR effect. The two different coercive fields of the layers H_{c1} and H_{c2} (dotted lines) and the resulting magnetic field H of the given input currents $+I$ and $-I$ are indicated in the upper panel.

step) used by new types of algorithms. (4) Aside from the input and output lines no further means are necessary for the logic operation, for example, additional magnetic fields, switching thresholds, or voltages. (5) The switching frequency of magnetic films can be pushed to several GHz, in principle allowing fast operation¹¹. (6) Since only short current pulses are required to rotate the magnetization, the energy consumption is further reduced. (7) The size of an MRAM cell—commercially available in 2004 (ref. 12)—can easily be scaled down to less than 100 nm (refs 13, 14) to further increase the integration density. (8) Our approach a priori does not necessarily require semiconductors to fabricate a logic gate. (9) The run-time programmability will increase the computational efficiency.

Our new approach has the potential to induce a paradigm shift from transistor-based logic to magneto-logic, where the programmable functionality and non-volatility are as important as miniaturization and clock speed. □

Received 12 February; accepted 1 September 2003; doi:10.1038/nature02014.

1. Prinz, G. A. Magnetoelectronics. *Science* **282**, 1660–1663 (1998).
2. Sidhu, R. P. S., Mei, A. & Prasanna, V. K. in *Field-programmable Logic and Applications* (eds Lysaght, P., Irvine, J. & Hartenstein, R. W.) 301–312 (Lecture Notes in Computer Science 1673, Springer, Berlin, 1999).
3. Parkin, S. S. P. et al. Exchange-biased magnetic tunnel junctions and application to nonvolatile magnetic random access memory. *J. Appl. Phys.* **85**, 5828–5833 (1999).
4. Grünberg, P. Layered magnetic structures: history, highlights, applications. *Phys. Today* **54**, 31–37 (2001).
5. Baibich, B. N. et al. Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
6. Binasch, G., Grünberg, P., Saurenbach, F. & Zinn, W. Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. *Phys. Rev. B* **39**, 4828–4830 (1989).
7. Moodera, J. S., Kinder, L. R., Wong, T. M. & Meservey, R. Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. *Phys. Rev. Lett.* **74**, 3273–3276 (1995).
8. Black, W. C. Jr & Das, B. Programmable logic using giant-magnetoresistance and spin-dependent tunneling devices. *J. Appl. Phys.* **87**, 6674–6679 (2000).
9. Richter, R. et al. Field programmable spin-logic based on magnetic tunneling elements. *J. Magn. Mater.* **240**, 127–129 (2002).
10. Martin, A. J. *Proceedings of SSGRR 2000, International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*; at (<http://www.ssgrr.it/en/ssgrr2000/papers/185.pdf>) (2000).
11. Gerrits, Th., van den Berg, H. A. M., Hohlfield, J., Bär, L. & Rasing, Th. Ultrafast precessional magnetization reversal by picosecond magnetic field pulse shaping. *Nature* **418**, 509–511 (2002).
12. Motorola Inc. Press release. IEEE International Solid State Circuits Conference (San Francisco, June 2001); at (<http://www.motorola.com/mot/document/content/0,1028,372,00.doc>) (2000).
13. Goronkin, H., von Allmen, P., Tsui, R. K. & Zhu, T. X. *Nanostructure Science and Technology* (eds Siegel, R. W., Hu, E. & Roco, M. C.) 67–92 (National Science and Technology Council (NSTC) Committee on Technology and The Interagency Working Group on NanoScience, Engineering and Technology (IWGN), 1999); at (http://www.wtec.org/loyola/nano/05_01.htm) (1999). Copyright is held by: WTEC, Loyola College (Maryland).
14. Compañ, R. *Technological Roadmap for European Nanoelectronics* at (<ftp://ftp.cordis.lu/pub/itn/docs/fetnldrm.zip>) (2000).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (koch@pdi-berlin.de).

Controlling anisotropic nanoparticle growth through plasmon excitation

Rongchao Jin, Y. Charles Cao, Encai Hao, Gabriella S. Métraux, George C. Schatz & Chad A. Mirkin

Department of Chemistry and Institute for Nanotechnology, Northwestern University, Evanston, Illinois 60208, USA

Inorganic nanoparticles exhibit size-dependent properties that are of interest for applications ranging from biosensing^{1–5} and catalysis⁶ to optics⁷ and data storage⁸. They are readily available in a wide variety of discrete compositions and sizes^{9–14}. Shape-selective synthesis strategies now also yield shapes other than

nanospheres, such as anisotropic metal nanostructures with interesting optical properties^{15–23}. Here we demonstrate that the previously described photoinduced method²³ for converting silver nanospheres into triangular silver nanocrystals—so-called nanoprisms—can be extended to synthesize relatively monodisperse nanoprisms with desired edge lengths in the 30–120 nm range. The particle growth process is controlled using dual-beam illumination of the nanoparticles, and appears to be driven by surface plasmon excitations. We find that, depending on the illumination wavelengths chosen, the plasmon excitations lead either to fusion of nanoprisms in an edge-selective manner or to the growth of the nanoprisms until they reach their light-controlled final size.

The photoinduced synthesis of silver (Ag) nanoprisms involves the preparation of a colloidal suspension of Ag spheres (diameter <10 nm), followed by conversion of the spheres to larger prism structures with visible light. In a typical experiment, colloidal Ag nanoparticles passivated with sodium citrate and bis(*p*-sulphonatophenyl)phenylphosphine dihydrate dipotassium (BSPP) (diameter 4.8 ± 1.1 nm, standing solution) were irradiated with a narrow-band light source (using a 150 W xenon lamp with a light output ~ 12 W) with an optical bandpass filter (centre wavelength 550 nm, width 40 nm) for ~ 50 h (see Supplementary Information). Transmission electron microscopy (TEM) shows that the colloid formed consists of two different size distributions of nanoprisms (Fig. 1a and inset), with the smaller particles (designated as type 1) and the larger particles (type 2) having average edge lengths of 70 ± 12 nm and 150 ± 16 nm, respectively. These structures tend to form stacks, so that edge-on views allow the precise determination of nanoprisms thickness²³. Although the average edge lengths for the type 1 and type 2 nanoprisms are significantly different, their thicknesses are almost identical (9.8 ± 1.0 nm) (Fig. 1b, c).

The bimodal particle growth process also has been monitored by ultraviolet–visible–near-infrared (UV–vis.–NIR) spectroscopy (Fig. 2a). During the reaction, one sees the disappearance of the plasmon band at ~ 395 nm (characteristic of the spherical silver particles) and the formation of two new, strong plasmon bands at 680 nm and 1,065 nm that are associated with the type 1 and type 2 nanoprisms, respectively (see below). The band for the type 1 prisms is initially centred at $\lambda_{\max} = 680$ nm and gradually blue-shifts to $\lambda_{\max} = 640$ nm. This blue-shifting correlates with the tip sharpness of the nanoprisms features; rounding is known to lead to blue-shifting²⁴. The second strong band at $\lambda_{\max} = 1,065$ nm is assigned to type 2 particles (see below). In addition to the two strong surface plasmon bands, one can observe two other weak resonances at 340 and 470 nm, respectively (Fig. 2a, spectrum 6). To gain further insight into the optical spectra of the solution with the bimodal particle distribution, we carried out theoretical modelling using a finite-element-based method known as the discrete dipole approximation (DDA)^{24–26}. The calculated spectrum shows plasmon bands that reproduce the experimentally observed spectrum (compare Fig. 2b and Fig. 2a, spectrum 6), confirming our peak assignments. The first three peaks in the spectrum of the colloid containing both type 1 and type 2 particles, centred at 340 nm (out-of-plane quadrupole resonance), 470 nm (in-plane quadrupole resonance) and 640 nm (in-plane dipole resonance)²⁴, result from the type 1 nanoprisms; in the case of the type 2 nanoprisms, only the strong dipole resonance at 1,065 nm is clearly observed. Quadrupole resonances, which occur at 340 nm and 600 nm (weak) in the spectrum of the solution of the type 2 nanoprisms, are overlapped with plasmon bands from the type 1 nanoprisms (Fig. 2b). The time-dependent optical spectra thus suggest that the process is bimodal, rather than unimodal as would be expected in the case of conventional Ostwald ripening^{13,14}.

The bimodal growth of Ag nanoprisms is not caused by the wavelength dispersity of the excitation beam (550 ± 20 nm). Indeed, when a monochromatic laser beam ($\lambda = 532.8$ nm, the

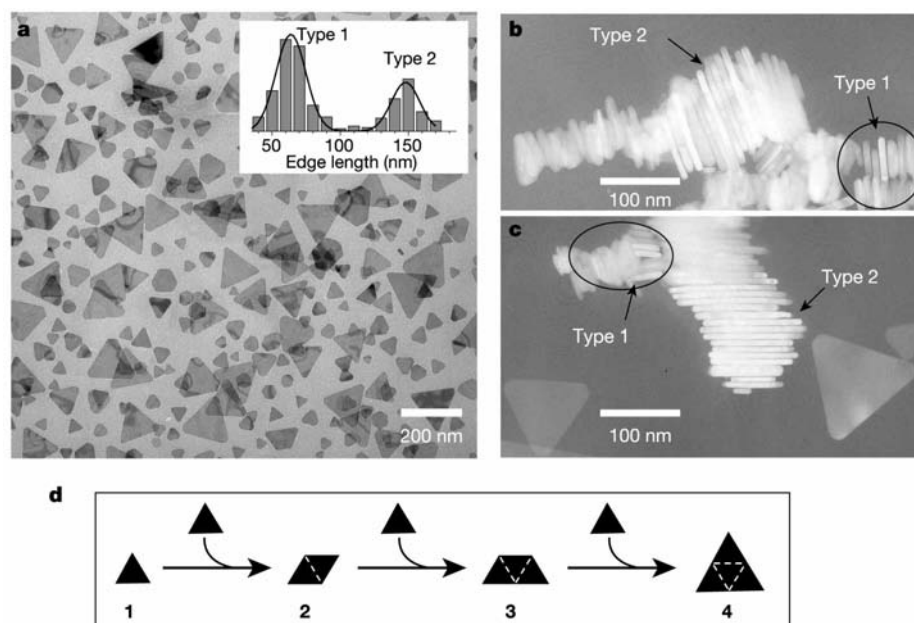


Figure 1 The bimodal growth of Ag nanoprisms. **a**, TEM image of a sample of Ag nanoprisms formed using single-beam excitation (550 ± 20 nm); inset, histograms used to characterize the size distribution as bimodal. **b**, **c**, TEM images of nanoprism stacks

showing that nanoprisms have nearly identical thicknesses (9.8 ± 1.0 nm). **d**, Schematic diagram of the proposed light-induced fusion growth of Ag nanoprisms.

second harmonic of a Nd:YAG laser, CW, light output ~ 0.2 W) is used to photolyse the Ag colloids, bimodal growth is still observed (see Supplementary Information). The bimodal growth was also observed with other excitation wavelengths (500–700 nm). Additional experiments where the surfactant BSPP was absent from the reaction mixture and initial Ag colloid yielded comparable results, demonstrating that BSPP is not critical for effecting the bimodal growth process and nanoprism formation. In addition, the bimodal growth process was observed for a variety of BSPP:sodium citrate ratios (investigated molar ratios ranged from 0:1 to 1:1, with sodium citrate fixed at 0.3 mM) and with different silver salt

precursors (AgNO_3 , $\text{CH}_3\text{CO}_2\text{Ag}$, AgClO_4 and Ag_2SO_4).

We propose that the observed bimodal growth process occurs through an edge-selective particle fusion mechanism, with four type 1 nanoprisms coming together in step-wise fashion to form a type 2 nanoprism (Fig. 1d). The following observations are consistent with this mechanism. First, bimodal growth results in type 1 and type 2 prisms where four of the former prisms can fit together to form a prism with dimensions (cumulative edge length, 140 ± 17 nm) that compare well with the latter (150 ± 16 nm) (Fig. 1). Second, edge-selective growth occurs with no apparent change in nanostructure thickness in going from the type 1 to type 2

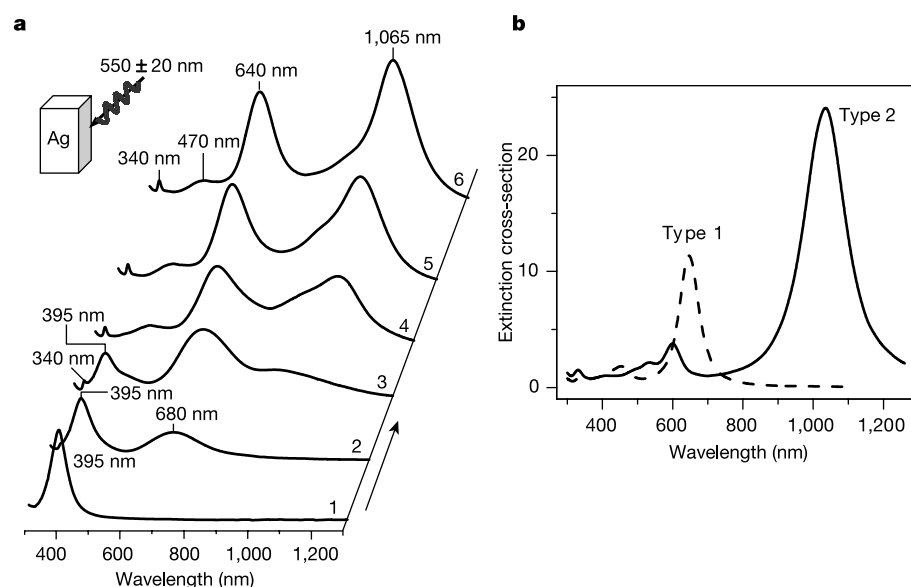


Figure 2 The optical spectra of Ag nanoprisms. **a**, Time evolution of UV-vis-NIR spectra of a Ag colloid (4.8 ± 1.1 nm spheres) under single-beam excitation (550 ± 20 nm). Spectrum 1, initial colloid; 2, after 10 h; 3, after 15 h; 4, after 19 h; 5, after 24 h; and 6, after 55 h. **b**, Theoretical modelling of the optical spectra of two different-sized

nanoprisms (model parameters: edge length of type 1, 70 nm, type 2, 150 nm; thickness, 10 nm). Note that the tip truncation of nanoprisms, which leads to a blueshift of the dipole resonance from 770 to 640 nm, has been taken into account in the modelling.

prisms. Third, detailed time-dependent UV–vis.–NIR measurements show that the onset of the growth of the band at 1,065 nm (assigned to type 2) is significantly delayed in comparison with the growth of the band at 640 nm (assigned to type 1) (Fig. 2a). This indicates that the fusion of nanoprisms occurs only after type 1 nanoprisms have accumulated. Fourth, a small population of dimer 2 and trimer 3 intermediates (Fig. 1d) is observed during the early stages of type 2 particle growth (see Supplementary Information).

We also performed electrostatics calculations on the optical properties of possible intermediate species involved in the fusion growth process. The results show that intermediates 2 and 3 have dipole plasmon excitations close to 600 and 1,065 nm (see Supplementary Information). Type 1 particles and the intermediates 2 and 3 can thus all absorb light at 600 nm, which can lead to the excited state needed for particle fusion to occur. However, the type 2 particles do not show dipole plasmon excitation explaining why they represent the end of the particle growth path. Note in this context that removal of surface ligands has, in the case of CdTe (ref. 27) and PbSe (C. B. Murray, personal communication), resulted in the fusion of spherical particles into nanowire structures; similar examples involving spherical particle fusion have also been reported²⁸.

At first glance, the observed bimodal growth appears to contradict previous results in which unimodal nanoprism growth was observed when visible light (white) from a conventional fluorescent tube was used as the excitation source²³. By careful analysis of the optical properties of these nanostructures and the effects of photolysis on them, we have identified a type of surface plasmon cooperativity in the photochemistry of Ag nanoprisms. To demonstrate this cooperative effect on nanoprism growth, we excited a solution of Ag nanoparticles (4.8 ± 1.1 nm) at two wavelengths, 550 ± 20 nm (primary) and 450 ± 5 nm (secondary) ($I_{550}:I_{450} = 2:1$, Fig. 3a). The 450-nm wavelength was selected to excite the quadrupole plasmon of the type 1 prisms. Double-beam excitation at these wavelengths inhibits the formation of type 2 nanoprisms

and results in exclusive formation of the smaller type 1 nanoprisms (72 ± 8 nm), as evidenced by UV–vis.–NIR spectra and TEM analysis (Fig. 3b, spectrum 4, and Fig. 3e). We further investigated the effect of varying the wavelength of the secondary beam and found that a 550-nm/340-nm coupled beam, in which the 340-nm light coincides with the out-of-plane quadrupole plasmon of the type 1 nanoprisms, also can inhibit the formation of type 2 nanoprisms and result in unimodal growth. However, in the cases of 550-nm/395-nm, 550-nm/610-nm, and 550-nm/650-nm coupled beams, in which the secondary wavelengths fall within the dipole resonances of the Ag nanospheres (395 nm) and type 1 nanoprisms (610 and 650 nm), respectively, bimodal growth is observed (see Supplementary Information). These data strongly indicate that only secondary wavelengths that can excite quadrupole plasmon modes can inhibit bimodal growth. Indeed, it is this photo-cooperativity that leads to the results observed with a fluorescent tube as the excitation source²³. The emission spectrum of a fluorescent tube exhibits bands at 546 nm and 440 nm, and has the appropriate intensity ratio (100%:40%) to effect photosynthetic cooperativity and hence unimodal growth. Consistent with this conclusion, when a 550 ± 20 nm band filter is used with a fluorescent tube to effect the photosynthetic conversion, bimodal growth is observed.

This observation of photo-cooperativity provides a way of controlling particle size with light. By supplementing the primary light source (450–700 nm) with a fixed secondary beam (340 nm, corresponding to out-of-plane quadrupole plasmon excitation), we can intentionally effect unimodal growth and generate a solution of nanoprisms of a desired average size. Using this approach, we have been able to synthesize nanoprisms with in-plane dipole plasmon resonances that track with particle size from 30 to 120 nm by using primary excitation wavelengths of 450 ± 20 nm, 490 ± 20 nm, 520 ± 20 nm, 550 ± 20 nm, 650 ± 20 nm and 750 ± 20 nm, respectively (Fig. 3b–f). The average edge lengths of the resulting nanoprisms correlate well with the wavelength of the primary

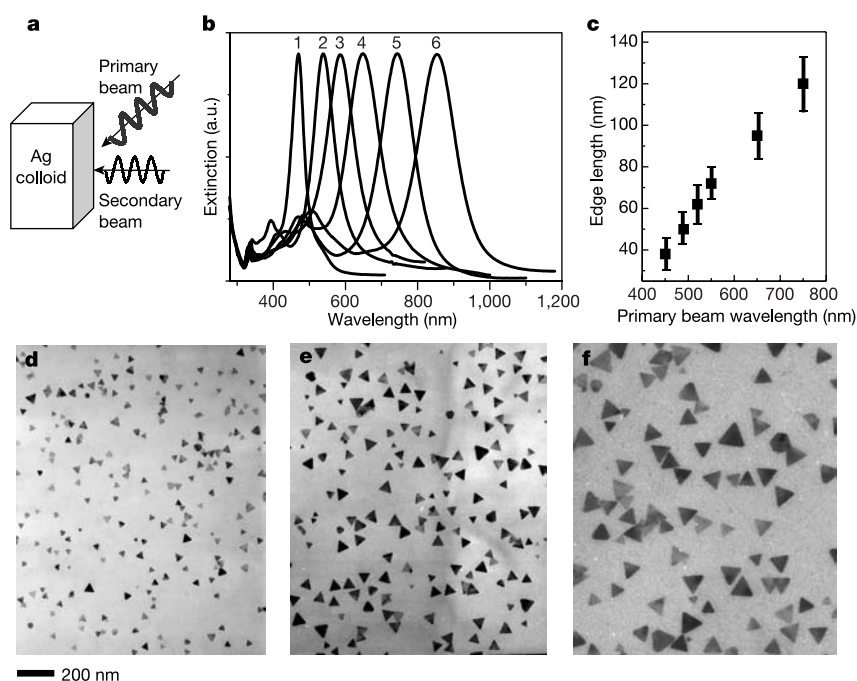


Figure 3 The unimodal growth of nanoprisms. **a**, Schematic diagram of dual-beam excitation. **b**, The optical spectra (normalized) for six different-sized nanoprisms (1–6 edge length: 38 ± 7 nm, 50 ± 7 nm, 62 ± 9 nm, 72 ± 8 nm, 95 ± 11 nm and 120 ± 14 nm) prepared by varying the primary excitation wavelength (central wavelength at 450, 490, 520, 550, 650 and 750 nm, respectively; width, 40 nm) coupled

with a secondary wavelength (340 nm; width, 10 nm). **c**, The edge lengths as a function of the primary excitation wavelength. **d–f**, TEM images of Ag nanoprisms with average edge lengths of 38 ± 7 nm (**d**), 72 ± 8 nm (**e**) and 120 ± 14 nm (**f**). Scale bar applies to panels **d–f**.

excitation source (Fig. 3c), which shows that a longer primary excitation wavelength produces larger particles with in-plane dipole plasmons (the red-most peak in each spectrum) that are red-shifted with respect to the excitation wavelength (Fig. 3b).

Another feature of using wavelength to control particle size is that subsequent addition of Ag spherical particles (4.8 ± 1.1 nm) to the nanoprisms does not lead to enlargement of the nanoprisms (see Supplementary Information); instead, the particles added photochemically grow into nanoprisms similar in size to the present ones (as determined by the excitation wavelength). This is in contrast to thermal methods for controlling particle sizes, in which addition of precursors typically leads to larger particles¹⁴. Note that the wavelength control of particle size is not likely to be a result of photothermal (or optical 'burning') effects; such effects have been invoked in other studies involving intense pulse laser irradiation of metal nanostructures (for example, 10^6 W)^{29,30}. The light source used to effect nanoprisms conversion is very weak (beam power ≤ 0.2 W). Indeed, according to the equation $\Delta T = \Delta H/C_p$ (where ΔH is the absorbed photon energy, and C_p is the heat capacity of silver, $0.235 \text{ J K}^{-1} \text{ g}^{-1}$), single 550-nm photon absorption by a type 1 prism can only lead to a negligible increase in temperature (≤ 0.007 K) (see Supplementary Information). The cumulative experimentally determined temperature increase after 50 h of photolysis (550 ± 20 nm) was less than 10°C .

Surface plasmons are typically studied as physical properties of metal nanostructures rather than chemical tools that provide control over growth and ultimate particle dimensions. The results reported here provide clear evidence for the importance of plasmon excitation in the Ag nanoprisms growth process, both for type 1 particles (which apparently grow from the initially produced colloidal particles to a size that depends on the dipole plasmon wavelength) and for type 2 particles (whose growth also requires dipole plasmon excitation, but is inhibited by quadrupole plasmon excitation). Although a detailed mechanism for these types of conversions remains to be determined, it is possible that plasmon excitation does two things. First, it could redistribute charge on the surfaces of the type 1 nanoprisms to either facilitate (in the case of dipole excitation) or inhibit (in the case of quadrupole excitation) particle-particle fusion. In addition, surface plasmon excitation could facilitate ligand dissociation at the particle edges (as this is where the local fields are the most intense²⁴), allowing the type 1 particles to grow through the addition of silver atoms or clusters. Taken together, these results are consistent with a new type of particle size control that is initiated and driven by light, highly cooperative, and surface-plasmon directed. □

Received 28 March; accepted 28 August 2003; doi:10.1038/nature02020.

- Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
- Bruchez, M., Moronne, M., Gin, P., Weiss, S. & Alivisatos, A. P. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
- Han, M., Gao, X., Su, J. Z. & Nie, S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nature Biotechnol.* **19**, 631–635 (2001).
- Nicewarner-Peña, S. R. et al. Submicrometer metallic barcodes. *Science* **294**, 137–141 (2001).
- Cao, Y. C., Jin, R. & Mirkin, C. A. Nanoparticles with Raman spectroscopic fingerprints for DNA and RNA detection. *Science* **297**, 1536–1540 (2002).
- Schmid, G. Large clusters and colloids. Metals in the embryonic state. *Chem. Rev.* **92**, 1709–1727 (1992).
- Wang, J. F., Gudiksen, M. S., Duan, X. F., Cui, Y. & Lieber, C. M. Highly polarized photoluminescence and photodetection from single indium phosphide nanowires. *Science* **293**, 1455–1457 (2001).
- Sun, S., Murray, C. B., Weller, D., Folks, L. & Moser, A. Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal superlattices. *Science* **287**, 1989–1992 (2000).
- Nirmal, M. et al. Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature* **383**, 802–804 (1996).
- Henglein, A. Radiolytic preparation of ultrafine colloidal gold particles in aqueous solution: optical spectrum, controlled growth, and some chemical reactions. *Langmuir* **15**, 6738–6744 (1999).
- Brust, M. & Kiely, C. J. Some recent advances in nanostructure preparation from gold and silver particles: a short topical review. *Colloid Surf. A* **202**, 175–186 (2002).
- Korgel, B. A. & Fitzmaurice, D. Self-assembly of silver nanocrystals into two-dimensional nanowire arrays. *Adv. Mater.* **10**, 661–665 (1998).
- Talapin, D. V. et al. Dynamic distribution of growth rates within the ensembles of colloidal II–VI and III–V semiconductor nanocrystals as a factor governing their photoluminescence efficiency. *J. Am. Chem. Soc.* **124**, 5782–5790 (2002).

- Peng, X., Wickham, J. & Alivisatos, A. P. Kinetics of II–VI and III–V colloidal semiconductor nanocrystal growth: "focusing" of size distributions. *J. Am. Chem. Soc.* **120**, 5343–5344 (1998).
- Chang, S. S., Shih, C. W., Chen, C. D., Lai, W. C. & Wang, C. R. C. The shape transition of gold nanorods. *Langmuir* **15**, 701–709 (1999).
- Sun, Y. G. & Xia, Y. N. Shape-controlled synthesis of gold and silver nanoparticles. *Science* **298**, 2176–2179 (2002).
- Maillard, M., Giorgio, S. & Pileni, M. P. Silver nanodisks. *Adv. Mater.* **14**, 1084–1086 (2002).
- Maillard, M., Giorgio, S. & Pileni, M. P. Tuning the size of silver nanodisks with similar aspect ratios: synthesis and optical properties. *J. Phys. Chem. B* **107**, 2466–2470 (2003).
- Pastoriza-Santos, I. & Liz-Marzan, L. M. Synthesis of silver nanoprisms in DMF. *Nano Lett.* **2**, 903–905 (2002).
- Chen, S. H., Fan, Z. Y. & Carroll, D. L. Silver nanodisks: synthesis, characterization, and self-assembly. *J. Phys. Chem. B* **106**, 10777–10781 (2002).
- Hao, E. C., Kelly, K. L., Hupp, J. T. & Schatz, G. C. Synthesis of silver nanodisks using polystyrene mesospheres as templates. *J. Am. Chem. Soc.* **124**, 15182–15183 (2002).
- Sun, Y. & Xia, Y. Triangular nanoplates of silver: synthesis, characterization, and use as sacrificial templates for generating triangular nanorings of gold. *Adv. Mater.* **15**, 695–699 (2003).
- Jin, R. et al. Photoinduced conversion of silver nanospheres to nanoprisms. *Science* **294**, 1901–1903 (2001).
- Kelly, K. L., Coronado, E., Zhao, L. L. & Schatz, G. C. The optical properties of metal nanoparticles: the influence of size, shape, and dielectric environment. *J. Phys. Chem. B* **107**, 668–677 (2003).
- Yang, W. H., Schatz, G. C. & Van Duyne, R. P. Discrete dipole approximation for calculating extinction and Raman intensities for small particles with arbitrary shapes. *J. Chem. Phys.* **103**, 869–875 (1995).
- Draine, B. T. & Flatau, P. J. Discrete-dipole approximation for scattering calculations. *J. Opt. Soc. Am. A* **11**, 1491–1499 (1994).
- Tang, Z., Kotov, N. A. & Giersig, M. Spontaneous organization of single CdTe nanoparticles into luminescent nanowires. *Science* **297**, 237–240 (2002).
- Penn, R. L. & Banfield, J. F. Imperfect oriented attachment: dislocation generation in defect-free nanocrystals. *Science* **281**, 969–971 (1998).
- Link, S., Burda, C., Mohamed, M. B., Nikoobakht, B. & El-Sayed, M. A. Laser photothermal melting and fragmentation of gold nanorods: energy and laser pulse-width dependence. *J. Phys. Chem. B* **103**, 1165–1170 (1999).
- Kamat, P. V. Photophysical, photochemical and photocatalytic aspects of metal nanoparticles. *J. Phys. Chem. B* **106**, 7729–7744 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We acknowledge the use of a Cary 500 spectrometer in the Keck Biophysics Facility at Northwestern University. C.A.M. and G.C.S. thank the AFOSR, ONR, DARPA and NSF for support of this work. R.J. is grateful for the support of the American Chemical Society Cognis Fellowship in Colloid and Surface Chemistry.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.A.M. (camirkin@chem.northwestern.edu) or G.C.S. (schatz@chem.northwestern.edu).

Enantiospecific electrodeposition of a chiral catalyst

Jay A. Switzer, Hiten M. Kothari, Philippe Poizot, Shuji Nakanishi & Eric W. Bohannon

Department of Chemistry and Graduate Center for Materials Research, University of Missouri-Rolla, Rolla, Missouri 65409-1170, USA

Many biomolecules are chiral—they can exist in one of two enantiomeric forms that only differ in that their structures are mirror images of each other. Because only one enantiomer tends to be physiologically active while the other is inactive or even toxic, drug compounds are increasingly produced in an enantiomerically pure form¹ using solution-phase homogeneous catalysts and enzymes. Chiral surfaces offer the possibility of developing heterogeneous enantioselective catalysts that can more readily be separated from the products and reused. In addition, such surfaces might serve as electrochemical sensors for chiral molecules. To date, chiral surfaces have been obtained by adsorbing chiral molecules^{2–6} or slicing single crystals so that they exhibit high-index faces^{7–13}, and some of these surfaces act as enantioselective heterogeneous catalysts^{5,6,10}. Here we show that

chiral surfaces can also be produced through electrodeposition, a relatively simple solution-based process that resembles biomineralization^{14–17} in that organic molecules adsorbed on surfaces have profound effects on the morphology of the inorganic deposits^{18–20}. When electrodepositing a copper oxide film on an achiral gold surface in the presence of tartrate ion in the deposition solution, the chirality of the ion determines the chirality of the deposited film, which in turn determines the film's enantiospecificity during subsequent electrochemical oxidation reactions.

There have been elegant experiments in which achiral surfaces are modified by chiral molecules to impart enantiospecificity to the surface^{2–5}. It has been shown, for instance, that tartaric acid adsorbed onto both Cu(110) and Ni(110) produces chiral surfaces^{2,3}. Cysteine adsorbed on Au(110) from a racemic mixture forms molecular pairs that are exclusively homochiral⁴. Raney nickel modified with (*R,R*)-tartaric acid can be used to catalyse the hydrogenation of β -ketoesters, producing the *R*-product with over 90% enantiomeric excess⁵. Switching the enantiomer of the adsorbate switches the product to the *S*-isomer. One problem with this approach to heterogeneous catalysis is that the adsorption of chiral modifiers needs to be carefully maintained during the synthesis⁶.

Another approach to the preparation of chiral heterogeneous catalysts is to use high-index surfaces of single crystals^{7–13}. These high-index surfaces are prepared by slicing a low-index single crystal at an angle. The high-index faces of face-centred cubic (f.c.c.) metals can exhibit chirality owing to kink sites on the surface. For example, Pt and Au metal crystals with (643) and ($\bar{6}\bar{4}\bar{3}$) faces are enantiomorphs. Only the surface of these materials is chiral, because the f.c.c. metals are highly symmetrical and do not have chiral space groups. The (531) surface of Pt has been shown to be enantioselective for the electrochemical oxidation of L-glucose¹⁰.

Our approach to the formation of stable surfaces that function as enantiospecific heterogeneous catalysts and sensors is to electrodeposit epitaxial films of low-symmetry materials such as monoclinic CuO onto high-symmetry achiral surfaces such as cubic Au(001). The chirality of solution precursors controls the handedness of the electrodeposited film. We have previously shown that electrodeposition can be used to deposit epitaxial films of metal oxides on single-crystal metal^{21–24} and semiconductor²⁵ surfaces. We have also found that the deposition solution and applied potential

can have a profound effect on the crystallographic orientation and morphology of the epitaxial films. For example, films of electrodeposited Cu₂O have a crystallographic orientation that is pH dependent. A film of Cu₂O deposited on Au(001) at pH 12 undergoes a transition from a thermodynamically controlled orientation to a kinetically controlled orientation after reaching a critical thickness²⁴.

The CuO films in this study were deposited using a method outlined previously²⁶. The electrodeposited CuO has a monoclinic structure (space group *C2/c*) with $a = 0.4685$ nm, $b = 0.3430$ nm, $c = 0.5139$ nm and $\beta = 99.08^\circ$. A Bragg–Brentano X-ray diffraction pattern is shown in Fig. 1 for an epitaxial film of CuO on Au(001) that was electrodeposited from a solution of Cu(II)(*S,S*)-tartrate. The film has a strong $[1\bar{1}\bar{1}]$ orientation, indicating that the system has a CuO($1\bar{1}\bar{1}$)/Au(001) epitaxial relationship. Other workers have shown that CuO with a $[111]$ orientation can be grown on MgO(001) by vapour deposition²⁷. In vapour deposition, however, the ultra-high-vacuum conditions preclude the use of solution precursors to control the film chirality. In the present work, the orientation of electrodeposited CuO can be changed to $[\bar{1}11]$ by depositing the film from a solution of Cu(II)(*R,R*)-tartrate. The $[1\bar{1}\bar{1}]$ and $[\bar{1}11]$ orientations are not distinguishable by Bragg–Brentano scans, because the *d*-spacings for the two orientations are identical.

The absolute configuration of the electrodeposited films was determined by X-ray pole figure analysis. X-ray diffraction determines the orientation of the entire film, not just the surface. Pole figures can be used to probe planes that are not parallel with the geometric surface of the sample. The sample is moved through a series of tilt angles, χ , and at each tilt angle the sample is rotated through azimuthal angles, ϕ , of 0° to 360° . Peaks occur in the pole figure when the Bragg condition is satisfied. Pole figures are shown in Fig. 2a and b for CuO films that were deposited from (*S,S*)- and (*R,R*)-tartrate solutions, respectively. The (111) planes of CuO were

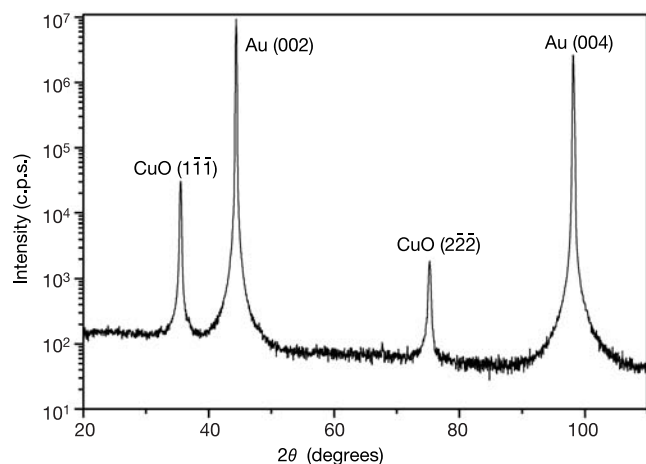


Figure 1 X-ray diffraction of electrodeposited CuO. Bragg–Brentano (θ – 2θ) X-ray diffraction scan probing the out-of-plane orientation of CuO that was electrodeposited from a solution of Cu(II)(*S,S*)-tartrate onto a single-crystal Au(001) surface. Only the $(1\bar{1}\bar{1})$ and $(2\bar{2}\bar{2})$ peaks of CuO are observed, indicating that the system has the CuO($1\bar{1}\bar{1}$)/Au(001) epitaxial relationship. The X-ray radiation is CuK α_1 , with a wavelength of 0.1540562 nm. c.p.s., counts per second.

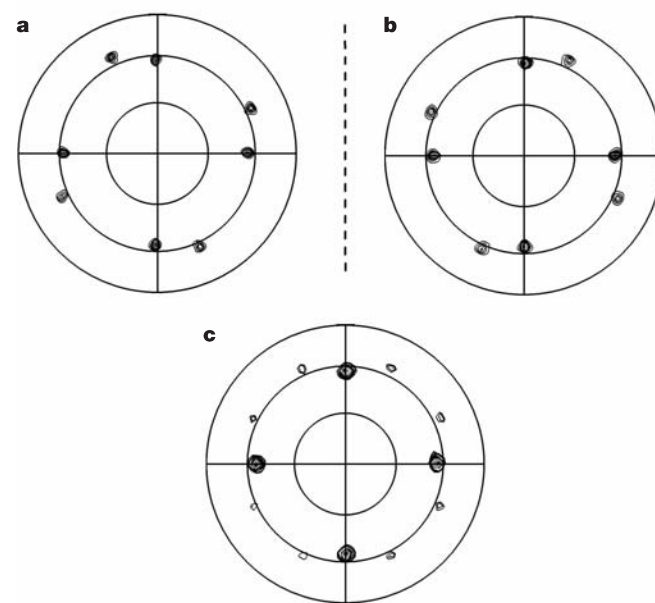


Figure 2 Determination of the absolute configuration of chiral CuO. X-ray pole figures of CuO films on Au(001) deposited from **a**, Cu(II)(*S,S*)-tartrate, **b**, Cu(II)(*R,R*)-tartrate, and **c**, racemic Cu(II)(tartrate). The film grown in Cu(II)(*S,S*)-tartrate has a $[1\bar{1}\bar{1}]$ orientation, and the film grown in Cu(II)(*R,R*)-tartrate has a $[\bar{1}11]$ orientation. The two films are enantiomorphs. The film in **a** has an *S* configuration, and the film in **b** has an *R* configuration. The film in **c** deposited from the racemic mixture shows equal amounts of the *R* and *S* configurations. The radial grid lines on the pole figures correspond to 30° increments of the tilt angle.

probed because they are close in d -spacing to those of Au(111). Therefore, there are four peaks at $\chi = 55^\circ$ that result from the Au. These serve as an internal reference point for the CuO peaks. Overlapping with the four Au peaks are peaks due to CuO($\bar{1}\bar{1}\bar{1}$) in Fig. 2a and CuO($\bar{1}\bar{1}\bar{1}$) in Fig. 2b. There are also four peaks at $\chi = 63^\circ$ that correspond to CuO($\bar{1}\bar{1}\bar{1}$) in Fig. 2a and CuO(111) in Fig. 2b. By comparison with stereographic projections for the monoclinic structure, these can be assigned as a $[\bar{1}\bar{1}\bar{1}]$ orientation for the film grown in (S,S)-tartrate (Fig. 2a), and a $[\bar{1}\bar{1}\bar{1}]$ orientation for the film grown in (R,R)-tartrate (Fig. 2b). In each case there are four equivalent in-plane orientations, with the $[110]$ direction of CuO coincident with the $[110]$, $[\bar{1}\bar{1}0]$, $[\bar{1}\bar{1}0]$ and $[\bar{1}\bar{1}0]$ directions of Au. The two pole figures in Fig. 2a and b are non-superimposable mirror images, indicating that the two films are enantiomers. Figure 2c is a pole figure of a CuO film deposited from a racemic mixture of (S,S)- and (R,R)-tartrate. This film has equal amounts of the two enantiomeric orientations.

The chiral deposition scheme is outlined in Fig. 3. The surfaces shown are ideal terminations of the bulk structure. In this figure, the smaller Cu atoms are shown dark red, and there are two distinct oxygen atoms. The filled, blue-coloured oxygen atoms are closest to the Cu plane, and sit in three-fold hollow sites. The open, blue-coloured oxygen atoms are situated nearly atop the Cu atoms. The $[\bar{1}\bar{1}\bar{1}]$ and $[\bar{1}\bar{1}\bar{1}]$ orientations of CuO shown in the figure are non-superimposable mirror images. Although CuO has an achiral space group, the $(\bar{1}\bar{1}\bar{1})$ and $(\bar{1}\bar{1}\bar{1})$ faces are enantiomorphs because they lack a centre of symmetry.

The handedness of the CuO films is determined by the chirality of the deposition solution, because the Au(001) surface has high symmetry and does not impart the chirality. We speculate that the chiral electrodeposition of CuO is directed by the adsorption of either free tartrate ions or Cu(II)(tartrate) complexes on the Au surface. Adsorption onto the initial Au(001) surface does seem to provide an essential imprint, because the electrodeposited CuO does not change orientation if the film is first grown in Cu(II)(R,R)-tartrate and then the deposition solution is switched to Cu(II)(S,S)-tartrate. Complexes of Cu(II)(tartrate) have a dimeric structure with a symmetry that is determined by the

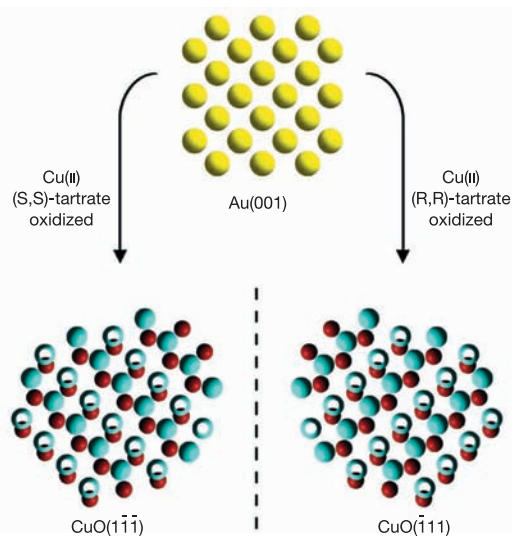


Figure 3 Outline of the enantiospecific electrodeposition scheme. Chiral CuO with either a $[\bar{1}\bar{1}\bar{1}]$ or $[\bar{1}\bar{1}\bar{1}]$ orientation is electrodeposited onto achiral Au(001). The smaller dark red spheres at the bottom of the figure represent Cu atoms. There are two non-equivalent O atoms which are coloured blue. The filled, blue-coloured O atoms are closest to the Cu plane, and sit in three-fold hollow sites. The open, blue-coloured O atoms are nearly atop the Cu atoms. The two orientations of CuO are clearly non-superimposable mirror images.

handedness of the tartrate ligand²⁸. Tartaric acid is known to adsorb on both Cu(110) and Ni(110) single crystals to form chiral surfaces^{2,3}. The reduction of symmetry of surfaces by the adsorption of chiral molecules is well known in biomineralization to produce chiral morphologies on minerals such as calcite and gypsum, which have achiral space groups^{14–17}. Selective adsorption on the surfaces of minerals such as calcite has been invoked to explain the genesis of biogenic homochirality²⁹.

Using the method of Attard¹², an R or S designation can be determined for the two enantiomorphs. By analogy to the Cahn–Ingold–Prelog sequence rules for organic molecules, an arbitrary ‘priority’ is assigned to each of the low-index planes of a crystal on the basis of the surface packing density. For f.c.c. metals this sequence is $\{111\} > \{100\} > \{110\}$. If the $\{111\} \rightarrow \{100\} \rightarrow \{110\}$ sequence runs clockwise in the stereographic projection of the material along a particular zone axis, the orientation is designated ‘ R ’. Anticlockwise rotation yields the designation ‘ S ’. Although this notation is arbitrary, it does allow one to assign a label to each of the enantiomers. In our work, we found by azimuthal X-ray diffraction that the R -enantiomer of CuO deposits with an 85% enantiomeric excess in the (R,R)-tartrate solution, and the S -enantiomer deposits with a 90% enantiomeric excess in the (S,S)-tartrate solution.

The pole figures show that the bulk films grown in (R,R)- and (S,S)-tartrate are enantiomers, but they do not provide information on the chirality of the surface. In order to probe the surface chirality, the electrochemical activity for films deposited in the two solutions

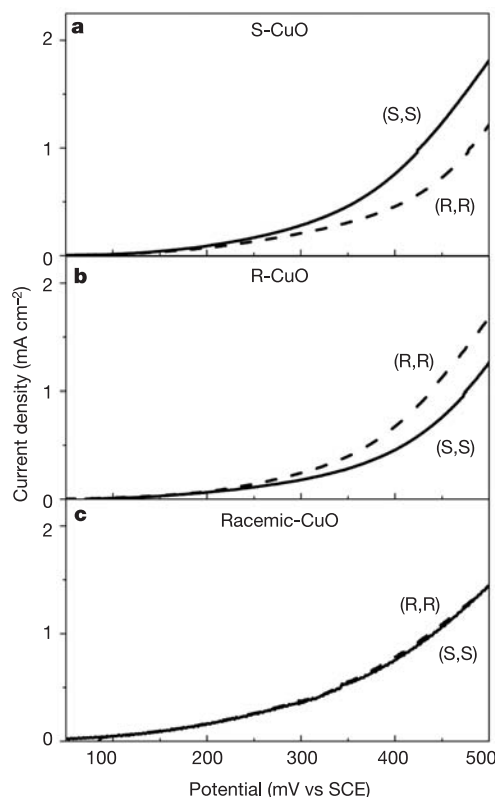


Figure 4 Chiral recognition by electrodeposited CuO. Linear sweep voltammograms comparing the electrocatalytic activity of **a**, an (S)-CuO film grown in Cu(II)(S,S)-tartrate with that of **b**, an (R)-CuO film grown in Cu(II)(R,R)-tartrate for the oxidation of tartrate. The (S)-CuO($\bar{1}\bar{1}\bar{1}$) film is enantioselective for the oxidation of (S,S)-tartrate, and the (R)-CuO($\bar{1}\bar{1}\bar{1}$) film is enantioselective for the oxidation of (R,R)-tartrate. A control film deposited from racemic Cu(II)(tartrate) shown in **c** has no enantioselectivity. The voltammograms were run at room temperature at a sweep rate of 10 mV s^{-1} in unstirred solutions of uncomplexed 5 mM (S,S)- and (R,R)-tartrate in 0.1 M NaOH. The (S,S)-tartrate and (R,R)-tartrate voltammograms are designated with solid and dashed lines, respectively. SCE, saturated calomel electrode.

was compared for the electrochemical oxidation of uncomplexed (S,S)- and (R,R)-tartrate. CuO has been shown by other workers to be a potent electrocatalyst for the oxidation of carbohydrates, amino acids, simple alcohols, aliphatic diols, and alkyl polyethoxy alcohol detergents³⁰. Chiral recognition by CuO has not, to our knowledge, been demonstrated previously. Linear sweep voltammograms are shown in Fig. 4 for the oxidation of (S,S)- and (R,R)-tartrate on CuO electrodes that were deposited from Cu(II)(S,S)-tartrate (Fig. 4a) and Cu(II)(R,R)-tartrate (Fig. 4b). The (S)-CuO film grown in (S,S)-tartrate is more active for the oxidation of the (S,S)-tartrate, and the (R)-CuO film grown in (R,R)-tartrate is more active for the oxidation of the (R,R)-tartrate. A control film shown in Fig. 4c (which was deposited from a racemic mixture of the (S,S)- and (R,R)-tartrates) shows no selectivity for the oxidation of the enantiomers.

We believe that our approach should be quite general for the deposition of other enantiospecific catalysts, because it does not require that the materials crystallize in a chiral space group. One attractive potential application of such catalysts would be their use as post-chromatographic chiral electrochemical sensors, which would no longer require chiral separation before chemical detection. □

Methods

Solutions, sample preparation, and electrochemistry

The CuO films were deposited to a thickness of approximately 300 nm at 30 °C onto a polished and H₂-flame-annealed Au(001) single crystal at an anodic current density of 1 mA cm⁻² from a solution of 0.2 M Cu(II), 0.2 M tartrate ion, and 3 M NaOH. Chiral recognition was demonstrated using uncomplexed tartrate ion as the target molecule. The linear sweep voltammograms in Fig. 4 were run at room temperature in unstirred solutions of 5 mM (S,S)- and (R,R)-tartrate in 0.1 M NaOH at a sweep rate of 10 mV s⁻¹.

X-ray diffraction

X-ray diffraction measurements were done on a high-resolution Philips X'Pert MRD diffractometer. For the Bragg-Brentano scan, the primary optics module was a combination Göbel mirror and a two-crystal Ge(220) two-bounce hybrid monochromator, and the secondary optics module was a 0.18° parallel plate collimator. The hybrid monochromator produces pure CuKα₁ radiation (λ = 0.1540562 nm) with a divergence of 25 arcsec. Pole figures were obtained in point-focus mode using a crossed-slit collimator as the primary optics, and a 0.27° parallel plate collimator and flat graphite monochromator as the secondary optics. A 2θ value of 38.742° was used to probe the (111) reflection of CuO. Enantiomeric excesses were determined from CuO(111) azimuthal scans at 2θ = 38.742° and χ = 63° by integrating the area under the (111) and (111̄) peaks due to the R and S forms of CuO, respectively.

Received 7 May; accepted 11 August 2003; doi:10.1038/nature01990.

- Stinson, S. C. Chiral pharmaceuticals. *Chem. Eng. News* **79**, 79–97 (2001).
- Lorenzo, M. O., Baddeley, C. J., Murryn, C. & Raval, R. Extended surface chirality from supramolecular assemblies of adsorbed chiral molecules. *Nature* **404**, 376–379 (2000).
- Humboldt, V., Haq, S., Murryn, C., Hofer, W. A. & Raval, R. From local adsorption stresses to chiral surfaces: (R,R)-tartaric acid on Ni(110). *J. Am. Chem. Soc.* **124**, 503–510 (2002).
- Kühnle, A., Linderer, T. R., Hammer, B. & Besenbacher, F. Chiral recognition in dimerization of adsorbed cysteine observed by scanning tunnelling microscopy. *Nature* **415**, 891–893 (2002).
- Izumi, Y. in *Advances in Catalysis* Vol. 32, 215–271 (Academic, New York, 1983).
- LeBlond, C., Wang, J., Liu, J., Andrews, A. T. & Sun, Y.-K. Highly enantioselective heterogeneously catalyzed hydrogenation of α-ketoesters under mild conditions. *J. Am. Chem. Soc.* **121**, 4920–4921 (1999).
- McFadden, C. E., Cremer, P. S. & Gellman, A. J. Adsorption of chiral alcohols on “chiral” metal surfaces. *Langmuir* **12**, 2483–2487 (1996).
- Horvath, J. D. & Gellman, A. J. Enantiospecific desorption of R- and S-propylene oxide from a chiral Cu(643) surface. *J. Am. Chem. Soc.* **123**, 7953–7954 (2001).
- Horvath, J. D. & Gellman, A. J. Enantiospecific desorption of chiral compounds from chiral Cu(643) and achiral Cu(111) surfaces. *J. Am. Chem. Soc.* **124**, 2384–2392 (2002).
- Ahmedi, A., Attard, G., Feliu, J. & Rodes, A. Surface reactivity at “chiral” platinum surfaces. *Langmuir* **15**, 2420–2424 (1999).
- Attard, G. A. *et al.* Temperature effects in the enantiomeric electro-oxidation of D- and L-glucose on Pt(643). *J. Phys. Chem. B* **103**, 1381–1385 (1999).
- Attard, G. A. Electrochemical studies of enantioselectivity of chiral metal surfaces. *J. Phys. Chem. B* **105**, 3158–3167 (2001).
- Sholl, D. S., Asthagiri, A. & Power, T. D. Naturally chiral metal surfaces as enantiospecific adsorbents. *J. Phys. Chem. B* **105**, 4771–4782 (2001).
- Orme, C. A. *et al.* Formation of chiral morphologies through selective binding of amino acids to calcite surface steps. *Nature* **411**, 775–779 (2001).
- Addadi, L. & Weiner, S. Crystals, asymmetry and life. *Nature* **411**, 753–755 (2001).
- Teng, H. H., Dove, P. M., Orme, C. A. & De Yoreo, J. J. Thermodynamics of calcite growth: baseline for understanding biomineral formation. *Science* **282**, 724–727 (1998).
- Cody, A. M. & Cody, R. D. Chiral habit modifications of gypsum from epitaxial-like adsorption of stereospecific growth inhibitors. *J. Cryst. Growth* **113**, 508–519 (1991).
- Paunovic, M. & Schlesinger, M. *Fundamentals of Electrochemical Deposition* 167–186 (Wiley-Interscience, New York, 1998).

- Josell, D., Wheeler, D., Huber, W. H. & Moffat, T. P. Superconformal electrodeposition in submicron features. *Phys. Rev. Lett.* **87**, 016102 (2001).
- Vazhev, V. V. & Mardashev, Yu. S. A “memory” effect in formic and oxalic acid oxidation at gold electrodes. *Sov. Electrochem.* **16**, 1445–1447 (1980).
- Switzer, J. A., Shumsky, M. G. & Bohannon, E. W. Electrodeposited ceramic single crystals. *Science* **284**, 293–296 (1999).
- Bohannon, E. W., Shumsky, M. G. & Switzer, J. A. Epitaxial electrodeposition of copper(I) oxide on single-crystal Au(100). *Chem. Mater.* **11**, 2289–2291 (1999).
- Sorenson, T. A., Morton, S. A., Waddill, G. D. & Switzer, J. A. Epitaxial electrodeposition of Fe₃O₄ thin films on the low-index planes of gold. *J. Am. Chem. Soc.* **124**, 7604–7609 (2002).
- Switzer, J. A., Kothari, H. M. & Bohannon, E. W. Thermodynamic to kinetic transition in epitaxial electrodeposition. *J. Phys. Chem. B* **106**, 4027–4031 (2002).
- Switzer, J. A., Liu, R., Bohannon, E. W. & Ernst, F. Epitaxial electrodeposition of a crystalline metal oxide onto single-crystalline silicon. *J. Phys. Chem. B* **106**, 12369–12372 (2002).
- Poizot, P., Hung, C.-J., Nikiforov, M. P., Bohannon, E. W. & Switzer, J. A. An electrochemical method for CuO thin film deposition from aqueous solution. *Electrochem. Solid-State Lett.* **6**, C21–C25 (2003).
- Catana, A., Locquet, J.-P., Paik, S. M. & Schuller, I. K. Local epitaxial growth of copper monoxide films on magnesium oxide. *Phys. Rev. B* **46**, 15477–15483 (1992).
- Missavage, R. J., Belford, R. L. & Paul, I. C. Crystal and molecular structure of tetrasodium dicopper(II)-dl-tartrate decahydrate. *J. Coord. Chem.* **2**, 145–157 (1972).
- Hazen, R. M., Filley, T. R. & Goodfriend, G. A. Selective adsorption of L- and D-amino acids on calcite: Implications for biochemical homochirality. *Proc. Natl Acad. Sci. USA* **98**, 5487–5490 (2001).
- Xie, Y. & Huber, C. O. Electrochemical detection and amperometric detection using an electrode made of copper oxide and carbon paste. *Anal. Chem.* **63**, 1714–1719 (1991).

Acknowledgements This work was supported by the National Science Foundation and the University of Missouri Research Board.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.S. (jswitzer@umr.edu).

Episodic sediment accumulation on Amazonian flood plains influenced by El Niño/Southern Oscillation

Rolf Aalto^{1,2}, Laurence Maurice-Bourgoin³, Thomas Dunne⁴, David R. Montgomery^{1,2}, Charles A. Nittrouer² & Jean-Loup Guyot⁵

¹Quaternary Research Center and ²Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA

³Institut de Recherche pour le Développement, UMR Laboratoire des Mécanismes de Transfert en Géologie, CP 7091 Lago Sul, CEP 71619-970, Brasília, DF, Brazil

⁴Donald Bren School of Environmental Science and Management and Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA

⁵IRD, UMR LMTG, CNRS/IRD Paul Sabatier University, F-31400 Toulouse, France

Continental-scale rivers with a sandy bed sequester a significant proportion of their sediment load in flood plains. The spatial extent and depths of such deposits have been described^{1,2}, and flood-plain accumulation has been determined at decadal time-scales^{3–5}, but it has not been possible to identify discrete events or to resolve deposition on near-annual timescales. Here we analyse ²¹⁰Pb activity profiles from sediment cores taken in the pristine Beni and Mamore river basins, which together comprise 720,000 km² of the Amazon basin, to investigate sediment accumulation patterns in the Andean–Amazonian foreland. We find that in most locations, sediment stratigraphy is dominated by discrete packages of sediments of uniform age, which are typically 20–80 cm thick, with system-wide recurrence intervals of about 8 yr, indicating relatively rare episodic deposition events. Ocean temperature and stream flow records link these episodic events to rapidly rising floods associated with La Niña events, which debouch extraordinary volumes of sediments from the Andes. We conclude that transient processes driven by the El

Niño/Southern Oscillation cycle control the formation of the Bolivian flood plains and modulate downstream delivery of sediments as well as associated carbon, nutrients and pollutants to the Amazon main stem.

Studies of modern flood-plain accumulation have described the spatial extent and depth of deposits^{1,2}, estimated flood-plain deposition using numerical models and surveys of grain size^{6–9}, or used a variety of mapping, field sampling and computations of the diffuse and channelized over-bank advection of turbid water¹⁰. Recent studies have measured flood-plain accumulation using geochronology sufficient to resolve these rates on decadal timescales^{3–5,11}. However, no previous study has resolved the spatial and temporal distribution of individual accumulation events across large, dynamic fluvial dispersal systems. It is now possible to identify discrete events and measure the accumulation from such processes with near-annual resolution^{12,13}. Our measurements along two rivers at a variety of distances from the channel reveal previously undocumented spatial and temporal patterns that reflect the hydraulics of sedimentation and the occurrence of flooding and sediment delivery forced by global atmospheric processes.

The Beni and Mamore rivers (Fig. 1) transport sediment from the

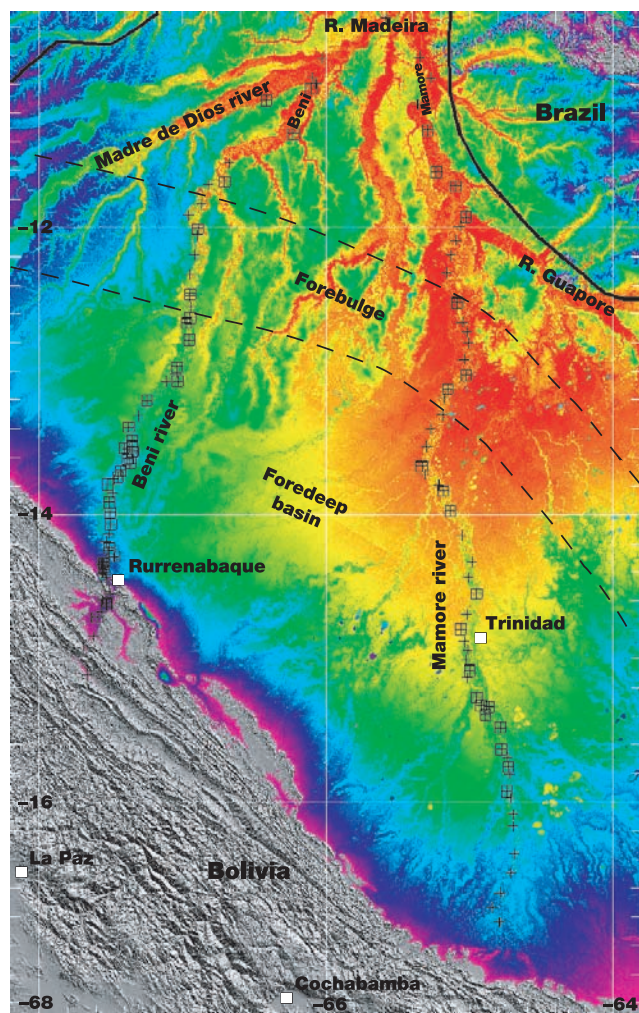


Figure 1 The Beni and Mamore river flood plains within the Llanos, northern Bolivia. The two rivers are the principal sediment and water sources for the Madeira River¹⁷, the largest sediment source for the Amazon¹⁰. Elevations >300 m are depicted as grey shaded relief; lower elevations as a rainbow equal-area stretch. Flood-plain transects are denoted with squares whereas in-channel samples are denoted with crosses. Approximate location of the foredeep and forebulge were determined with GPS surveys of longitudinal river gradient. One-degree latitude is approximately 110 km.

rapidly eroding Bolivian Andes¹⁴ across the flood plains of a large foreland basin^{15,16}. Flood plains and channels are essentially pristine, without artificial levees, dams, dredging, roads, significant deforestation or cultivation, or other anthropogenic complications. The Beni drains 70,000 km² of the northern Bolivian Andes into the lowland Amazon basin and then through an additional 50,000 km² of forested flood plain, depositing approximately 100 Mt yr⁻¹ of sediment as it traverses the foreland basin^{13,17,18}. The adjacent Mamore drains a 600,000-km² basin. These basins are representative of the vast expanse of Andean–Amazonian foreland basins to the north.

Sampling was conducted at locations representing a variety of channel–flood-plain geometries, during August and September of 1999 and 2000 (ref. 13). A total of 276 sediment cores 65–160-cm deep were extracted (153 from Beni and 123 from Mamore) from surveyed transects at distances of 50–1,200 m from the channel, predominately from the higher, mature flood plain on the cut-bank side of the river. Such spatial coverage allows resolution of discrete deposition events across transects traversing hundreds of metres of flood plain, with cores typically spaced every 50 m, as well as measurement of sedimentation along thousands of kilometres of river.

Cores were X-rayed to evaluate sedimentary structures and potential post-deposition disturbance. After imaging, 115 of the cores (80 from Beni and 35 from Mamore) were cut, processed, and clay-normalized ²¹⁰Pb activity profiles were measured with methods described and calibrated elsewhere^{12,19}, which allow the identification of individual sediment packages and their dating with annual resolution. The CIRCAUS (constant initial reach clay activity, unknown sedimentation) procedure¹² for flood-plain geochronology accounts for the basin-wide variation in the ²¹⁰Pb activity of river-borne sediment and relies on only a few assumptions, each of which is testable.

More than 95% of the Beni and Mamore cores depict episodic sediment accumulation best described by the CIRCAUS model¹². In most flood-plain locations, sediment arrives as discrete packages of uniform age across an observed depth range (Fig. 2). Granulometry and X-ray radiographs indicate that this sediment is silt-sized and is deposited as coherent packages in a low-energy environment. Fine horizontal laminations and other submillimetre-scale structures show that the flood-plain sediment has not been significantly disturbed since deposition and that bioturbation is negligible at most locations. Cross-bedding and other evidence for higher-energy deposition environments are rare on the cut-bank side of the flood plain, and are found mainly in point bar deposits close to the edge of the vegetation. Sediment packages are typically 20–80 cm thick, and vertical profiles from 6-m-high cut-bank outcrops indicate that the layers do not exceed 2 m in thickness. Hiatuses of decades occur between layers in cores with several events. Furthermore, the sediment at the flood-plain surface (after accounting for the meteoric cap) is often decades old, and in such cases is not blanketed by fresh sediment with higher ²¹⁰Pb activity from recent sedimentation events. Although Beni and Mamore flood-plain sediment arrives as discrete pulses, any particular flood-plain location receives sediment infrequently. Hence, episodic sedimentation is the predominant mechanism for flood-plain accumulation.

Averaged over these events, the accumulation rate varies with flood-plain distance from the active river channel (Fig. 3). High, variable rates proximal (<300 m) to the channel represent local processes that construct natural levees through frequent decanting of sediment over bank during annual floods—the few cores exhibiting constant sedimentation were located within this zone. Farther from the channel, lower, spatially uniform, temporally episodic rates reflect the extensive processes that convey most sediment onto the distal flood plain.

When accumulation dates from all cores are compiled (Fig. 4a), a basin-wide pattern emerges: distinct sedimentation pulses are

separated by years to decades. This century-long geochronological record spans the entire foreland, suggesting that floods needed to facilitate widespread accumulation occur approximately every 8 yr (11 events in 90 yr). This is a separate phenomenon from the previous observation that many decades may pass between sedimentation events at a specific flood-plain location. Because younger events bury older ones and most of the cores are about 1-m deep, our record is biased towards recent events and densely sampled river reaches. Therefore, only the dates of the peak centres should be considered, not the relative heights.

To interpret this temporal pattern, we investigated climatic variability. Eastern Bolivia experiences high rainfall in the Andes during cold-phase El Niño/Southern Oscillation (ENSO) events (La Niña)²⁰, causing the Beni river to flood²¹. We used a sea temperature index (STI) appropriate for the Beni river (Fig. 4a). Over the last century, the nine La Niña years match the sediment accumulation record. Two

exceptions, 1983 and 1998, are minor cold-phase ENSO years that follow intense warm-phase ENSO years. Sediment accumulation correlates well with such transitional cold-phase years (1975 is non-transitional La Niña, and 1977 the only exception).

The Mamore flood plain spans a transitional region of climate response to ENSO²², which seems to have migrated in the early 1970s: before 1970, rainfall was high during La Niña; after 1970, rainfall was lower during La Niña and higher or unaffected during El Niño²³. However, the rainfall response in the Andean tributaries of the Mamore has remained the same; that is, elevated La Niña rainfall. After 1970, Mamore flood-plain runoff was no longer synchronous with Andean sources, thus affecting flooding hydrology. For locations spanning the Mamore foredeep, no sediment has accumulated since 1974. This suggests a hiatus in significant flooding following a change in rainfall response to ENSO.

To explore the mechanisms governing flood-plain accumulation, we examined water discharge at Rurrenabaque (Fig. 4b). Sediment accumulation consistently occurs whenever discharge rises more than $8,000 \text{ m}^3 \text{ s}^{-1}$ over a few days. These rapid-rise years also correspond to peak discharges $>12,000 \text{ m}^3 \text{ s}^{-1}$. However, there are four slow-rise years with maximum discharges above that threshold, although these events are not recorded in our cores. As a result, the record of maximum flooding does not correlate to La Niña as well as does our flood-plain accumulation record.

Beni flood-plain topographic surveys indicate that flood-plain inundation begins when discharges exceed $6,000 \text{ m}^3 \text{ s}^{-1}$, a value exceeded annually¹⁸. However, distal flood plains do not exhibit chronic accumulation. Sedimentation is episodic in several specific ways: (1) events correspond only to large, rapid-rise floods; (2) neighbouring cores (within 1 km) typically record the same event, but there is little synchrony between locations tens of kilometres apart; and (3) at any specific location, sediment accumulates only during some floods, with a hiatus of decades typically separating sediment pulses 20–80 cm in thickness. The environment of deposition is low energy and silt rich. Suspended sediment in the Beni river is dominated by fine silts, whereas the bed is composed of sand^{13,24}, so source material for flood-plain sedimentation is derived from high in the water column. Inundation of the entire flood plain by sediment-laden water slowly decanted over levees during annual floods—the local mechanism that constructs the levees—would result in chronic, thin, synchronous sediment accumulation.

Instead, the thick episodic deposits are probably relatively small crevasse splays; delta-shaped deposits formed during levee failure. The banks are composed of loamy deposits susceptible to incision of

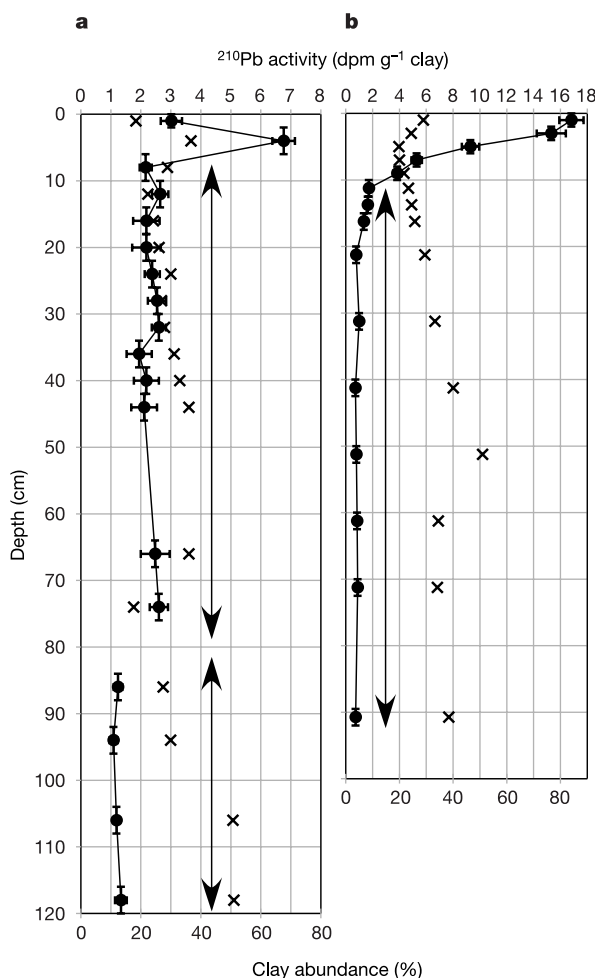


Figure 2 Unsupported ^{210}Pb activity profiles (circles) and clay abundance (crosses) from representative flood-plain cores. Accumulation events (denoted with vertical arrows) are dated with CIRCAUS geochronology^{12,13}. Increases within the top 10 cm represent meteoric fallout. **a**, Beni river site located 2,000 m and 2,900 m from the channel when the sediment was deposited in 1974 (± 1.8 yr) and 1951 (± 1.9 yr). The meteoric cap is incomplete. X-radiographs depict fine laminations throughout the upper core, with minor cross-bedding below 80 cm depth. **b**, Mamore river site located 175 m from the river when sediment was deposited in 1944 (± 2.0 yr). The cap date is 1945. X-radiographs depict fine laminations throughout the core. These examples portray typical episodic accumulation events, without chronic sedimentation. Nearby sites along flood-plain transects record similar events, a characteristic result that suggests that such sediment packages blanket regions hundreds to thousands of metres in width and length. dpm, disintegrations per minute.

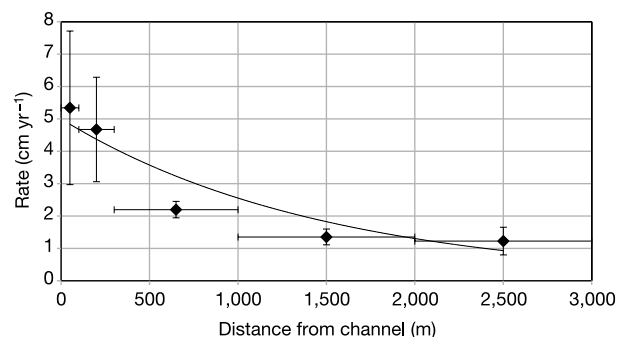


Figure 3 Spatial distribution of Beni flood-plain accumulation rates, averaged for five distance ranges, with best exponential fit ($r^2 = 0.82$). Because many flood-plain cores do not resolve the lower bound of accumulation events, rates are conservative. Owing to rapid channel migration across the flood plain, many of the core sites lay farther from the river during the most recent recorded deposition—concurrent river distances were measured from images (Landsat or aerial) bracketing the recorded date of each accumulation event. Projected across the flood plains of the Beni foreland basin¹⁸, this relationship provides for an estimated average net sediment loss of about 100 Mt yr^{-1} .

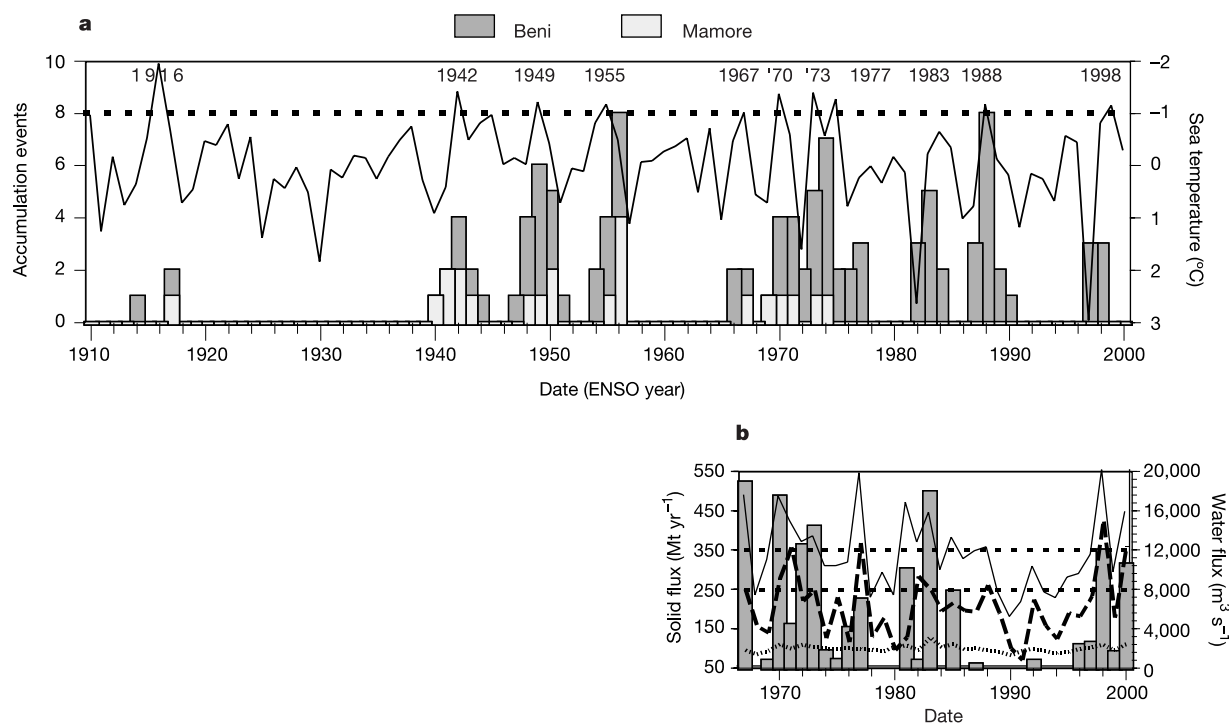


Figure 4 Temporal distribution of fluvial processes. **a**, Flood-plain accumulation events. STI is the average sea surface temperature anomaly in the eastern equatorial Pacific (150–80°W, 4°N–4°S) from November to February (the early rainy season months for Bolivia; temperatures are derived from a standard century-scale oceanographic record³⁰). Marked years exhibit either STI < −1° (La Niña, threshold marked) or the flood-plain accumulation of sediment (CIRCAUS geochronology cannot discriminate events 1 yr apart). ENSO, sediment and water years run from October to September. **b**, Maximum (solid curve) and mean annual (dotted curve) water discharge, and rise to flood peak (dashed, 2–3-day increase in discharge) discharged from the Andes (Beni river at

Rurrenabaque). The lower threshold shows both the minimum rapid flood rise (about 8,000 m³ s^{−1}) associated with flood-plain accumulation and the mean annual sediment flux (about 250 Mt). The upper threshold represents the minimum discharge associated with flood-plain accumulation (approximately 12,000 m³ s^{−1}). Also plotted is total sediment discharge (bars) conveyed by transitory bank-full floods (>6,000 m³ s^{−1})¹⁸, which comprise >60% of the average efflux—therefore, floods supply most of the sediment to the fluvial dispersal system and exceptional floods far surpass the mean annual flux and also account for much of the sediment exchanged due to channel migration¹⁸.

crevasse channels during rapid overtopping¹³. Floodwater, decanted from high in the water column, would diverge into the densely vegetated flood plain, depositing an extensive lens of fine sediment within this low-energy environment. Such deposits are similar to short-lived crevasse splays described in the literature^{2,25}, including ancestral foredeep deposits preserved in the Andes²⁶. Crevasse failure is more likely when the surrounding flood plain is relatively low, because of the higher hydraulic head or slope²⁷. Conversely, a high flood plain is typically inundated by sediment-free water²⁸ from: (1) local rainfall and tributaries; (2) river water that enters the flood plain through distant crevasses but deposits its sediment as it traverses the inundated forest; and (3) water that slowly decants over bank, depositing its sediment as it crosses the vegetated levees. The prevalence of such 'black' water inhibits crevasse formation by reducing head. Hence, only a rapid-rise flood can engender the critical elevation differential between water in the channel and in the flood plain required to form crevasses.

Crevasse splays, triggered by large, rapid-rise ENSO floods, account for the preponderance of flood-plain sediment accumulation across the Bolivian Llanos (lowland plains), building stratigraphic sequences that are discontinuous in both time and space. Such critical floods also dominate sediment discharge from Andean tributaries across the foreland and into the large rivers of the Amazonian lowland (Fig. 4b), indicating considerable inter-annual variation of sediment supply and transport resulting from the interaction of Andean erosion and the dynamics of ENSO-driven climate. However, annual water discharge is relatively constant and previous studies have interpreted the fluxes along the Amazon main stem and the resulting sedimentation processes as following a regular annual cycle^{10,29}. This implies an intervening region of active

sediment storage, exchange and geomorphic complexity. Because sediment supply, transport, deposition and exchange within these major Andean tributaries are orchestrated by extreme climate, such an 'alluvial capacitor' located between the foreland basins and the main-stem Amazon may further modulate particle transport to provide a more regular supply to Earth's largest river. □

Received 27 March; accepted 15 August 2003; doi:10.1038/nature02002.

1. Kesel, R. H., Yodis, E. G. & McCraw, D. J. An approximation of the sediment budget of the lower Mississippi River prior to major human modification. *Earth Surf. Process. Landforms* **17**, 711–722 (1992).
2. Aslan, A. & Autin, W. J. Evolution of the Holocene Mississippi River floodplain, Ferriday, Louisiana: insights of the origin of fine-grained floodplains. *J. Sedim. Res.* **69**, 800–815 (1999).
3. Nicholas, A. P. & Walling, D. E. Investigating spatial patterns of medium-term overbank sedimentation on floodplains: a combined numerical modeling and radiocesium-based approach. *Geomorphology* **19**, 133–150 (1997).
4. He, Q. & Walling, D. E. Use of fallout Pb-210 measurements to investigate longer-term rates and patterns of overbank sediment deposition on the floodplains of lowland rivers. *Earth Surf. Process. Landforms* **21**, 141–154 (1996).
5. Goodbred, S. L. Jr & Kuehl, S. A. Floodplain processes in the Bengal Basin and the storage of Ganges-Brahmaputra river sediment: an accretion study using Cs-137 and Pb-210 geochronology. *Sediment. Geol.* **121**, 239–258 (1998).
6. Mertes, L. A. K. Rates of flood-plain sedimentation on the central Amazon River. *Geology* **22**, 171–174 (1994).
7. Howard, A. D. in *Lowland Floodplain Rivers: Geomorphological Perspectives* (eds Carling, P. A. & Petts, G. E.) (Wiley and Sons, Chichester, UK, 1992).
8. Pizzuto, J. E. Sediment diffusion during overbank flows. *Sedimentology* **34**, 301–317 (1987).
9. James, C. S. Sediment transfer to overbank sections. *J. Hydraul. Res.* **23**, 435–452 (1985).
10. Dunne, T., Mertes, L. A. K., Meade, R. H., Richey, J. E. & Forsberg, B. R. Exchanges of sediment between the floodplain and channel of the Amazon River in Brazil. *Geol. Soc. Am. Bull.* **110**, 450–467 (1998).
11. Allison, M. A., Kuehl, S. A., Martin, T. C. & Hassan, A. Importance of flood-plain sedimentation for river sediment budgets and terrigenous input to the oceans: Insights from the Brahmaputra-Jamuna River. *Geology* **26**, 175–178 (1998).
12. Aalto, R. & Nittroer, C. A. Application of fallout ²¹⁰Pb geochronology to river-floodplain systems. *Sediment. Geol.* (submitted).
13. Aalto, R. *Geomorphic Form and Process of Mass Flux Within an Active Orogen: Denudation of the*

Bolivian Andes and Sediment Transport and Deposition Within the Channel-Floodplain Systems of the Amazonian Foreland. 365 Thesis, Univ. Washington (2002).

14. Aalto, R., Dunne, T. & Guyot, J. L. Geomorphic controls on Andean denudation rates. *J. Geol.* (in the press).
15. Horton, B. K. & DeCelles, P. G. The modern foreland basin system adjacent to the Central Andes. *Geology* **25**, 895–898 (1997).
16. Baby, P., Rochat, P., Mascle, G. & Hérail, G. Neogene shortening contribution to crustal thickening in the back arc of the Central Andes. *Geology* **25**, 883–886 (1997).
17. Guyot, J. L. *Hydrogéochimie des Fleuves de l'Amazonie Bolivienne*. Collection Études & Thèses (ORSTOM Ed.), Paris, (1993).
18. Aalto, R., Dunne, T., Maurice-Bourgoin, L., Guyot, J. L. & Nittrouer, C. A. Beni River morphology, migration, and sediment exchange between the river and its floodplain. *Geol. Soc. Am. Bull.* (submitted).
19. Nie, Y., Suayah, I. B., Benninger, L. K. & Alperin, M. J. Modeling detailed sedimentary 210-Pb and fallout 239–240-Pu profiles to allow episodic events: an application to Chesapeake Bay. *Limnol. Oceanogr.* **46**, 1425–1437 (2001).
20. Aceituno, P. On the functioning of the Southern Oscillation in the South America Sector Part I: Surface climate. *Mon. Weath. Rev.* **116**, 505–524 (1988).
21. Maurice-Bourgoin, L., Ronchail, J., Vauchel, P., Aalto, R. & Guyot, J. L. Climate controls (ENSO) on the flooding of the Beni River, a large Andean tributary of the Amazon. *EOS AGU Trans. Spring* A-09895 (2003).
22. Coelho, C. A. S., Uvo, C. B. & Ambrizzi, T. Exploring the impacts of the tropical Pacific SST on the precipitation patterns over South America during ENSO periods. *Theor. Appl. Climatol.* **71**, 185–197 (2002).
23. Ronchail, J. Variabilité pluviométrique en Bolivie lors des phases extrêmes de l'oscillation australe du Pacifique (1950–1993). *Bull. Inst. Fr. Etud. Andines* **27**, 687–698 (1998).
24. Guyot, J. L., Jouanneau, J. M. & Wasson, J. G. Characterisation of river bed and suspended sediments in the Rio Madeira drainage basin (Bolivian Amazonia). *J. Soc. Am. Earth Sci.* **12**, 401–410 (1999).
25. Smith, N. D., Cross, T. A., Dufficy, J. P. & Clough, S. R. Anatomy of an avulsion. *Sedimentology* **36**, 1–23 (1989).
26. Horton, B. K., Hampton, B. A. & Waanders, G. L. Paleogene synorogenic sedimentation in the Altiplano plateau and implications for initial mountain building in the central Andes. *Geol. Soc. Am. Bull.* **113**, 1387–1400 (2001).
27. Slingerland, R. L. & Smith, N. D. Necessary conditions for a meandering river avulsion. *Geology* **26**, 435–438 (1998).
28. Mertes, L. A. K. Description and significance of the perirheic zone on inundated floodplains. *Wat. Resour. Res.* **33**, 1749–1762 (1997).
29. Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M. & Hess, L. L. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO₂. *Nature* **416**, 617–620 (2002).
30. Meyers, S. D., Thelin, E. & O'Brien, J. J. Reconstruction of monthly SST in the tropical Pacific Ocean during 1868–1993 using adaptive climate basis functions. *Mon. Weath. Rev.* **127**, 1599–1612 (1999).

Acknowledgements This work was supported by an NSF research grant, a NASA Earth Systems Science Graduate Fellowship to R.A., and by the research collaboration (HYBAM Project) between IRD, SENAMHI (Bolivia), and the Universidad Mayor de San Andres, Bolivia. Laboratory assistance was provided by J. Staly, K. Sauers, G. Smith, J. Nittrouer and C. Gardner. Suggestions from C. Paola and A. Aufdenkampe improved the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A. (aalto@geomorphology.com).

Boundary-layer mantle flow under the Dead Sea transform fault inferred from seismic anisotropy

Georg Rümpler^{1,2}, Trond Ryberg¹, Günter Bock¹ & Desert Seismology Group*

¹GeoForschungsZentrum Potsdam, Telegrafenberg, 14473 Potsdam, Germany

* A complete list of authors appears at the end of the paper

Lithospheric-scale transform faults play an important role in the dynamics of global plate motion. Near-surface deformation fields for such faults are relatively well documented by satellite geodesy, strain measurements and earthquake source studies^{1,2}, and deeper crustal structure has been imaged by seismic profiling³. Relatively little is known, however, about deformation taking

place in the subcrustal lithosphere—that is, the width and depth of the region associated with the deformation, the transition between deformed and undeformed lithosphere and the interaction between lithospheric and asthenospheric mantle flow at the plate boundary. Here we present evidence for a narrow, approximately 20-km-wide, subcrustal anisotropic zone of fault-parallel mineral alignment beneath the Dead Sea transform, obtained from an inversion of shear-wave splitting observations along a dense receiver profile. The geometry of this zone and the contrast between distinct anisotropic domains suggest subhorizontal mantle flow within a vertical boundary layer that extends through the entire lithosphere and accommodates the transform motion between the African and Arabian plates within this relatively narrow zone.

At the southern end of the Dead Sea transform (DST), between the Dead Sea and the Red Sea, the Wadi Arava fault is the main active strike-slip fault^{4–6} trending approximately N20E. Near the

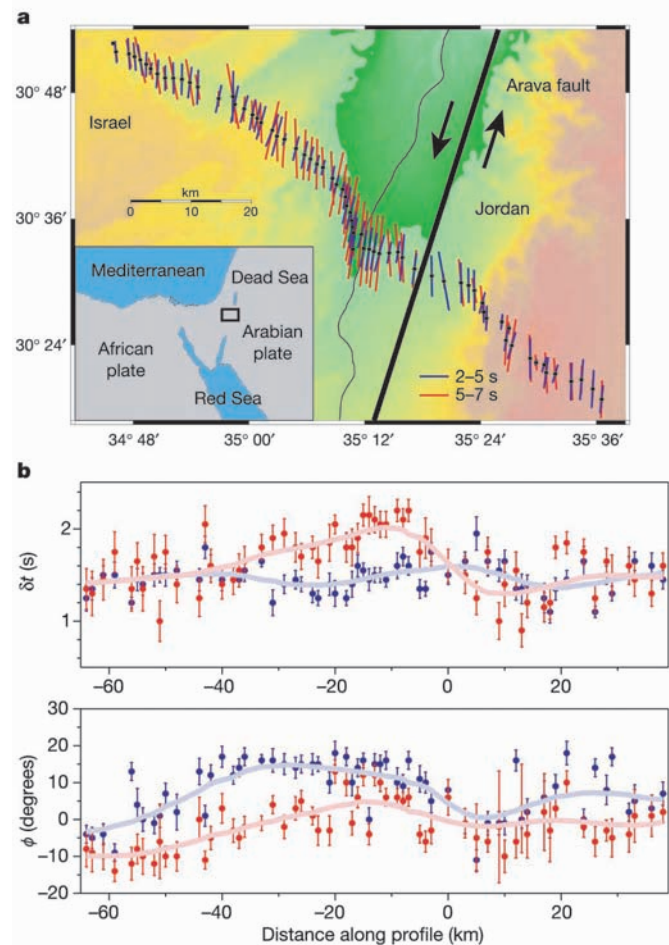


Figure 1 Study area and shear-wave splitting parameters. **a**, Map with topography and the locations of seismic stations for which the shear-wave splitting analysis was performed. The Arava fault (black line) strikes at approximately N20E. The bars indicate measured SKS splitting parameters for the period range of 2–5 s (blue) and 5–7 s (red). The orientation corresponds to the polarization direction of the fast shear wave (fast polarization ϕ) and the length is proportional to the delay time $\delta t - 1$ s. **b**, Measured shear-wave splitting parameters (circles) along the profile (delay times and fast polarization directions) for the two period bands (2–5 and 5–7 s). The parameters are obtained by application of an inverse splitting operator to minimize the energy of the transverse SKS component¹⁷. A measure of error has been derived from the 95% confidence region as determined by the χ^2 distribution. For reasons of representation a factor of 0.4 is applied to the error scales used in this figure. The lines represent a smoothed version of the measurements, calculated by averaging the results within a sliding window.

surface it is characterized by sinistral strike-slip motion of 105 km which has accumulated over the last 17 to 21 Myr (refs 7–10). Although there is evidence for normal faulting and east–west extension¹⁰, transcurrent motion clearly dominates. In the mantle, large-scale seismic anisotropy is caused by the deformation-induced preferred alignment of olivine crystals. Results from laboratory experiments¹¹ and numerical simulations of flow-induced textures¹² suggest that the olivine crystallographic *a*-axis aligns subparallel to the strike of transform-type plate boundaries. Away from the fault zone, the lithosphere is expected to retain its original crystallographic fabric frozen-in during previous tectonic events¹³. The analysis of shear-wave splitting of teleseismic phases (SKS) provides a tool to infer mineral texture and crystal alignment in the lithosphere. When traversing an anisotropic region, the incident shear wave is split into two phases that propagate at different wave speeds with orthogonal polarizations. The effect is quantified by the delay time (δt) between the two shear waves and by the polarization direction of the fast shear wave (ϕ). For a vertically travelling SKS phase, the fast polarization aligns parallel to a subhorizontal *a*-axis

orientation. Variations of the two splitting parameters with frequency and/or initial (source-induced) polarization direction are indicative of inhomogeneous anisotropy along the path of propagation¹⁴.

As part of the DESERT experiment¹⁵, 86 short-period (1 Hz) three-component seismometers were deployed along a 100-km profile crossing the DST about 75 km south of the Dead Sea basin (Fig. 1a). During the one-month recording period, a magnitude 7.6 (M_w) earthquake from the Volcano Island region was recorded with a high signal-to-noise ratio of the SKS phase for periods between 2 and 7 s. The rotation of the horizontal SKS waveform components into a radial-transverse coordinate system yields significant transverse-component energy (Fig. 2a, b), which is indicative of shear-wave splitting due to anisotropy beneath the profile¹⁶. The largest transverse amplitudes are found within an approximately 35-km-wide zone west of the surface location of the Arava fault. In the context of previous studies on SKS splitting, the relative transverse/radial amplitudes observed here are unusually large. However, for the period range considered (2–7 s), we show that such observations can be attributed solely to anisotropy, without the need to consider

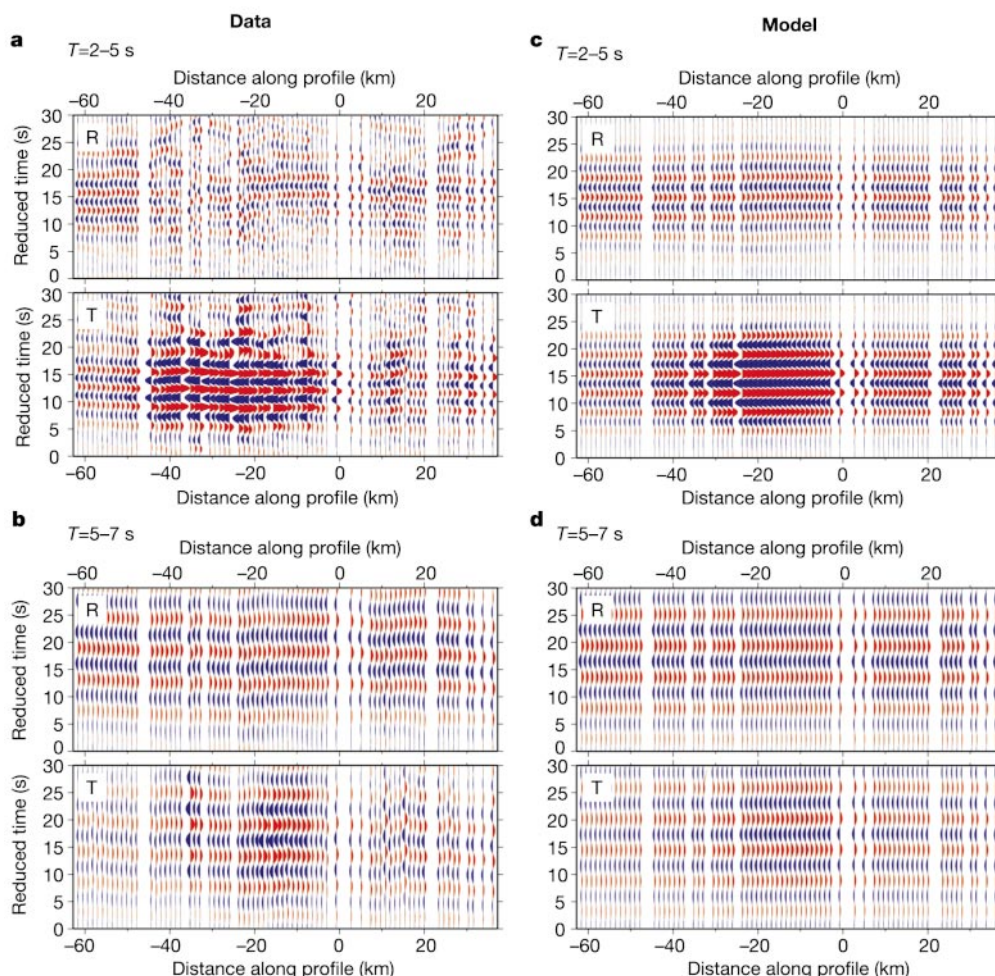


Figure 2 Radial (R) and transverse (T) components of SKS waveforms from the Volcano Island earthquake. (Earthquake origin time: 28 March 2000, 11:00:22.5; hypocentre coordinates: 22.34°N, 143.73°E, depth: 127 km; epicentral distance $\Delta = 94^\circ$, back-azimuth 61° with respect to the centre of the profile.) The amplitude scale is identical for all traces (radial and transverse), allowing a direct comparison of amplitudes. The presence of energy on the transverse component is indicative of anisotropy beneath the stations¹⁶. The traces are shown with respect to the distance along the profile, where the origin corresponds to the surface location of the Arava fault. **a**, A bandpass filter for periods (*T*) between 2 and 5 s has been applied to the seismograms. **b**, Seismograms after application of a bandpass filter between 5 and 7 s. **c**, **d**, The corresponding synthetic

SKS seismograms for the best-fitting model given in Fig. 4, calculated using a complete finite-difference method. The same bandpass filters have been applied to the data and the calculated seismograms. Generally, observed and calculated waveforms exhibit the same lateral and frequency-dependent variations. Some differences occur near -30 km, where the model leads to somewhat smaller *T* amplitudes at shorter periods. Also, at longer periods, observed *T* amplitudes between -35 and 0 km are slightly larger. The agreement between modelling and observations is expressed quantitatively by the variance reduction of 61%. Remaining differences between splitting parameters obtained from observed and calculated waveforms can possibly be attributed to deviations from the 2D geometry, or heterogeneous (isotropic) structures.

complexities related to three-dimensional (3D) heterogeneous structures. From the seismograms we obtained splitting parameters by application of an inverse operator that accounts for the effect of a hypothetical homogeneous anisotropic layer beneath the stations¹⁷. We determined splitting parameters for two period bands (1) 2 to 5 s and (2) 5 to 7 s (Supplementary Fig. S1). In the shorter-period range (2–5 s), delay times are relatively uniform with values of about 1.5 s. The polarization directions of the fast shear wave vary between N5W (−5°) and N15E (+15°) along the profile (Fig. 1b). Consistently larger values of ϕ , close to N15E, are found within the 35-km-wide zone that coincides with the largest transverse waveform components. In the longer-period range (5–7 s), delay times in the central region (to the west of the Arava fault) are consistently larger by 0.3 to 0.5 s. The fast polarizations remain close to north (0°) and differ by about 15° relative to the shorter-period results.

The coherent nature of the SKS waveforms recorded along the dense profile across the DST facilitates a simultaneous inversion of the observed wavefield for all stations. We use a local optimization technique to iteratively improve the fit between splitting parameters from synthetic and observed waveforms (see Supplementary Information). The starting model for the inversion is defined by the best-fitting model with a uniform anisotropic layer (Supplementary Fig. S2a), where the a -axis is aligned at 4° and 4.7% anisotropy. The variance σ between observed and calculated splitting parameters for this model is set to 1. From this, a laterally homogeneous two-layer model exhibiting an anisotropic crust and mantle is derived by the inversion (Supplementary Fig. S2b). We have not put any a priori constraints on the relative amount of crustal versus mantle anisotropy. Anisotropic parameters for this model are mainly constrained by the frequency dependence of the observed splitting parameters. Possible causes for significant anisotropy in the crust are stress-aligned cracks¹⁸ and a distinct mineral fabric in the ductile lower crust. This laterally homogeneous two-layer structure forms the starting model for the following inversions. In view of the three-part SKS amplitude variations along the profile, a first model to explain the observations is assumed to exhibit a uniformly anisotropic mantle, whereas the symmetry axis of the crustal anisotropy and its strength are assumed to vary laterally (Supplementary Fig. S2c). While the number of anisotropic blocks is kept fixed, their lateral extent is also determined during the inversion. To assess the possible significance of variations in mantle anisotropy, we consider a second model consisting of three distinct anisotropic domains in the mantle, whereas the crust is now uniformly anisotropic (Supplementary Fig. S2d). This type of model gives a significantly better fit to the observed splitting parameters, especially to the polarizations and their frequency dependence.

On the basis of results obtained for the latter two models, we further consider the combined effects of lateral variations of anisotropy in the crust and mantle. As indicated by the reduction of the relative variance between observed and calculated waveforms, the inversion favours those models that are characterized by relatively complex structured anisotropy in the mantle (Fig. 3). However, models consisting of more than three lateral domains in the mantle (or crust) are not considered, as smaller-scale variations of anisotropy cannot be resolved conclusively from the observations. The best-fitting model (Fig. 4) leads to a variance reduction of 61% compared to the reference model with a single uniformly anisotropic layer (Supplementary Fig. S2a). In the mantle, its anisotropy is characterized by a -axis directions of 3° beneath western stations and 11° beneath the eastern highlands. The central zone with an a -axis of 24° (parallel to the strike of the Arava fault) is about 20-km wide, and exhibits significantly stronger anisotropy, almost identical to the model with a uniform crust (Supplementary Fig. S2d). This highly anisotropic mantle zone with its a -axis subparallel to the surface trend of the DST suggests subhorizontal mantle flow within a narrow, approximately 20-km-wide, boundary layer that accommodates the lithospheric deformation between the

Arabian and African plates. Such a vertical decoupling zone is in agreement with recent results from thermomechanical modelling of the DST¹⁹ that predict a 20–40-km-wide zone, where shear strain localizes and lithospheric strength is minimal.

The modelling shows that the data are consistent with sharp changes of anisotropy in the mantle. The corresponding effects on the waveforms are smoothed out by diffraction and wavefront-healing (Fig. 2c, d). Much smoother variations of the model structure would result in further broadening of waveform variations, which is not consistent with the data. This requires that the transition between the decoupling zone and the surrounding mantle is significantly smaller than the dominant wavelength of the SKS phase in the medium, which suggests a transition of not more than 5-km width between distinct anisotropic domains. The maximum depth range for the anisotropic mantle zone is difficult to assess because an increase of its thickness, while simultaneously decreasing the strength of the anisotropy, has a relatively small effect on the wavefield calculations. We cannot rule out the possibility of a separate asthenospheric layer with a mineral fabric controlled by the absolute plate motion. Observations of fast polarizations from the Arabian shield show little variation²⁰, which may indicate the contribution from a regional-scale uniform mantle flow. However, accounting for this by including a uniform asthenospheric layer in the model leads to lateral variations between anisotropic domains in the lithosphere similar to those shown in Fig. 4 (Supplementary Fig. S5b).

The crustal anisotropy for the best-fitting model varies between 5% and 7% with symmetry axes from −21° through to −7° to −13°. The central and eastern crustal domains exhibit relatively small variations in symmetry-axis direction (by 6°), whereas differences in the percentage of anisotropy are more significant. This could indicate that the crust can be more-or-less characterized by two

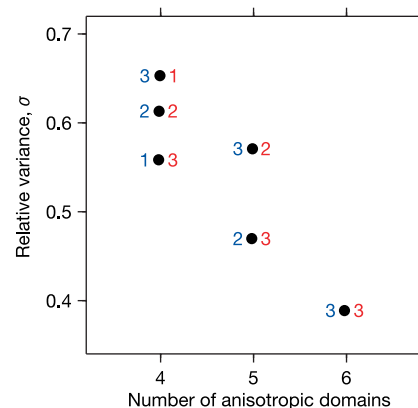


Figure 3 The relative variance (filled circles) between splitting parameters obtained from observed and calculated waveforms as function of the total number of anisotropic domains in the crust and mantle. Numbers in blue indicate the number of crustal domains for a given model; red numbers correspond to domains in the mantle. By increasing the total number of domains, the misfit between observed and calculated splitting parameters can be reduced further. However, the misfit for the four-block structure with three anisotropic domains in the mantle is smaller than the misfit for the (more complex) five-block structure with two anisotropic domains in the mantle. This result suggests complex anisotropic variations in the mantle. More than three lateral anisotropic domains cannot be resolved conclusively from the observations. (See Supplementary Figs S3 and S4 for parameters describing the models.) The models on which the inversions are based consist of a crust of uniform thickness (35 km) and a 100-km-thick mantle layer representing the lithosphere and the upper part of the asthenosphere. The thickness of the mantle layer is constrained from the globally observed average delay-time in continental regions of about 1 s, and under the initial assumption that the upper mantle exhibits anisotropy of about 4.5% (ref. 27). Variations in anisotropy are accounted for by two-dimensional (2D) block structures, each characterized by the orientation of the horizontal symmetry axis and the strength of the anisotropy (in per cent).

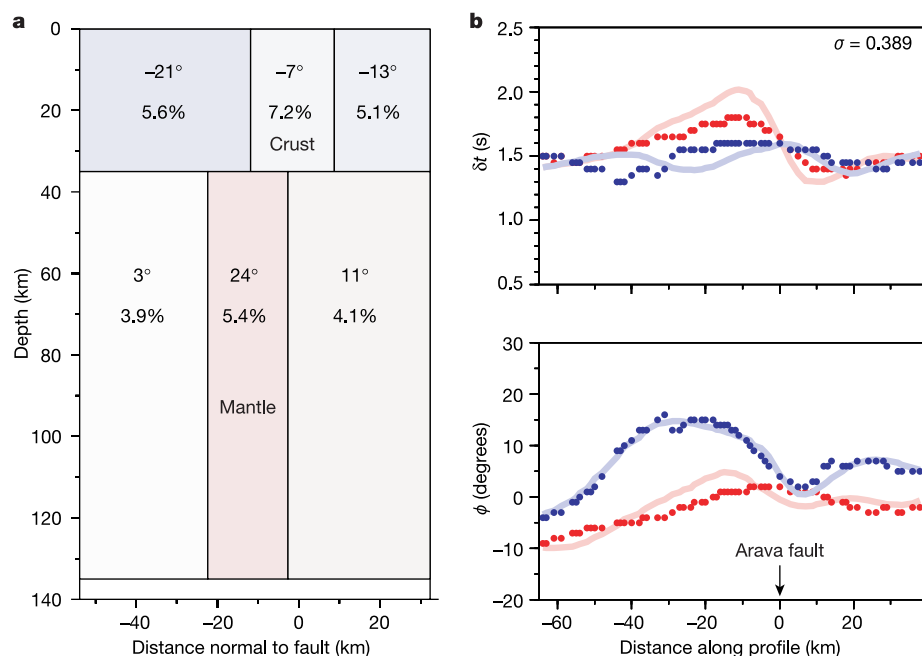


Figure 4 Model of anisotropy beneath the profile and comparison between observed and calculated shear-wave splitting parameters. **a**, Best-fitting 2D model with lateral variations of crustal and mantle anisotropy beneath the DST obtained from the simultaneous inversion of the observed splitting parameters. Angular values indicate the orientation of the horizontal symmetry axes within the anisotropic domains. The strength of the anisotropy (in per cent) is determined from the relative velocity difference between vertically propagating split shear waves. The effective isotropic shear-wave velocities are (crust) $v_s = 3.5 \text{ km s}^{-1}$ and (mantle) $v_s = 4.5 \text{ km s}^{-1}$, respectively. The central zone in the mantle with fault-parallel olivine *a*-axis and increased anisotropy is interpreted as a

vertical boundary layer of subhorizontal mantle flow that accommodates the transform motion between the Arabian and African plates. The model is shown with respect to distance normal to the Arava fault, which strikes at N20E. **b**, Lines represent (smoothed) splitting parameters obtained from the observed waveforms. Dots indicate corresponding splitting parameters (delay times and fast polarization directions) obtained from synthetic waveform modelling using a complete finite-difference method. The relative variance between observed and calculated splitting parameters is $\sigma = 0.39$, which corresponds to a variance reduction of 61% compared to the model with a single uniformly anisotropic layer (Supplementary Fig. S2a).

major units (Supplementary Fig. S4b). Interestingly, the boundary between these two structures would lie centred above the boundary layer in the mantle, but slightly to the west of the Arava fault. The crustal symmetry-axis orientations are compatible with the direction of maximum horizontal compression in the region²¹, and are similar to those predicted by modelling of the transform-related stress field¹⁹. The more highly anisotropic central domain may indicate a stronger contribution from stress-aligned cracks under the Arava valley, possibly suggesting a relatively broad region of intense faulting. Evidence for this comes from a number of faults to the west of the Arava fault, which have been revealed by near-surface seismic profiles and strike parallel to it²². Some of these faults may have been more active in the past, which would suggest a rather broad zone of crustal deformation not specifically constrained to a single fault. The Arava fault would thus mark the eastern boundary of this zone of deformation. Further modelling shows that purely isotropic velocity variations in the crust cannot account for the observed splitting parameters (Supplementary Fig. S5c).

Our shear-wave splitting results differ from those obtained at other major strike-slip faults. At stations in close proximity to the San Andreas fault the measurements suggest two-layer anisotropy in the mantle, where the upper layer exhibits a fault-parallel *a* axis, whereas the symmetry axis in the lower layer is oblique at about 45° to the fault. The crustal anisotropy seems to be negligible. Estimates for the width of the fault-zone-related anisotropy range from 50 to 150 km (refs 17, 23). In comparison, this could possibly indicate that the extent of the decoupling zone scales with the total strain accumulated along the fault. However, further observations from several large strike-slip faults within the India-Asian collision zone^{24–26} show a continuous rotation of fast polarization across the faults, indicating distinct anisotropy of neighbouring tectonic units. There, a boundary layer may be thinner or not developed

owing to the more pronounced compressional component of deformation. Also, the larger spacing between neighbouring stations may have precluded the detection of such a layer. □

Received 16 May; accepted 5 August 2003; doi:10.1038/nature01982.

- Holt, W. E., Bingming, S.-T., Haines, J. & Jackson, J. in *The History and Dynamics of Global Plate Motions* (eds Richards, M. A., Gordon, R. G. & van der Hilst, R. D.) 113–141 (American Geophysical Union, Washington DC, 2000).
- Shachak, P. *et al.* Current plate motion across the Dead Sea Fault from three years of continuous GPS monitoring. *Geophys. Res. Lett.* **29**, 10.1029/2001GL013879 (2002).
- Stern, T. A. & McBride, J. H. Seismic exploration of continental strike-slip zones. *Tectonophysics* **286**, 63–78 (1998).
- Ben-Menahem, A., Nur, A. & Vered, M. Tectonics, seismicity and structure of the Afro-Eurasian junction — the breaking of an incoherent plate. *Phys. Earth Planet. Inter.* **12**, 1–50 (1976).
- Garfunkel, Z. Internal structure of the Dead Sea leaky transform (rift) in relation to plate kinematics. *Tectonophysics* **80**, 81–108 (1981).
- Van Eck, T. & Hofstetter, A. Microearthquake activity in the Dead Sea region. *Geophys. J. Int.* **99**, 605–620 (1989).
- Quennel, A. M. The structural and geomorphic evolution of the Dead Sea rift. *Q. J. Geol. Soc. Lond.* **114**, 2–24 (1958).
- Freund, R. *et al.* The shear along the Dead Sea rift. *Phil. Trans. R. Soc. Lond. A* **267**, 107–130 (1970).
- McKenzie, D. P., Davies, D. & Molnar, P. Plate tectonics of the Red Sea and East Africa. *Nature* **224**, 125–133 (1970).
- Mart, Y. & Rabinowitz, P. The northern Red Sea and the Dead Sea Rift. *Tectonophysics* **124**, 85–113 (1986).
- Zhang, S. & Karato, S. Lattice preferred orientation of olivine aggregates deformed in simple shear. *Nature* **375**, 774–777 (1995).
- Tommasi, A., Tikoff, B. & Vauchez, A. Upper mantle tectonics: Three-dimensional deformation, olivine crystallographic fabrics and seismic properties. *Earth Planet. Sci. Lett.* **168**, 173–186 (1999).
- Silver, P. G. & Chan, W. W. Implications for continental structure and evolution from seismic anisotropy. *Nature* **335**, 34–39 (1988).
- Savage, M. K. & Marson-Pidgeon, K. Frequency dependent anisotropy in Wellington, New Zealand. *Geophys. Res. Lett.* **24**, 3289–3300 (1997).
- Desert Team, Multinational geoscientific research kicks off in the Middle East. *Eos* **81**, 609, 616–617 (2000).
- Kind, R., Kosarev, G. L., Makeyeva, L. I. & Vinnik, L. P. Observations of laterally inhomogeneous anisotropy in the continental lithosphere. *Nature* **318**, 358–361 (1985).
- Silver, P. G. Seismic anisotropy beneath the continents: Probing the depths of geology. *Annu. Rev. Earth Planet. Sci.* **24**, 385–432 (1996).

18. Crampin, S. Geological and industrial implications of extensive-dilatancy anisotropy. *Nature* **328**, 491–496 (1987).
19. Sobolev, S. V., Babeyko, A. Y. & Garfunkel, Z. Desert Group. Thermo-mechanical model of the Dead Sea Transform. *Earth Planet. Sci. Lett.* (submitted).
20. Wolfe, C. J., Vernon, F. L. III & Al-Amri, A. Shear-wave splitting across western Saudi Arabia: The pattern of upper mantle anisotropy at a Proterozoic shield. *Geophys. Res. Lett.* **26**, 779–782 (1999).
21. Reinecker, J., Heidbach O. & Müller, B. The 2003 release of the World Stress Map (<http://www.world-stress-map.org/>) (2003).
22. Bartov, Y., Avni, Y., Calvo, R. & Frieslander, U. The Zofar fault — A major intra-rift feature in the Arava Rift Valley. *Geol. Surv. Israel Curr. Res.* **11**, 27–32 (1998).
23. Savage, M. K. Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting? *Rev. Geophys.* **37**, 65–106 (1999).
24. McNamara, D. E., Owens, T. J., Silver, P. G. & Wu, F. T. Shear wave anisotropy beneath the Tibetan Plateau. *J. Geophys. Res.* **99**, 13655–13665 (1994).
25. Hirn, A. *et al.* Seismic anisotropy as an indicator of mantle flow beneath the Himalayas and Tibet. *Nature* **375**, 571–574 (1995).
26. Herquel, G., Tapponnier, P., Wittlinger, G., Mei, J. & Danian, S. Teleseismic shear wave splitting and lithospheric anisotropy beneath and across the Altyn Tagh fault. *Geophys. Res. Lett.* **26**, 3225–3228 (1999).
27. Nicolas, A. & Christensen, N. I. in *Composition, Structure and Dynamics of the Lithosphere-Asthenosphere System* (eds Fuchs, K. & Froidevaux, C.) *Geodyn. Ser.* **16**, 111–123 (1987).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to the Geophysical Institute of Israel, the National Ministry of Infrastructure of Israel, the Natural Resources Authority of Jordan, and the An-Najah University in Nablus, Palestine Authority, for their support. The instruments were provided by the Geophysical Instrument Pool of the GeoForschungsZentrum Potsdam. The experiment was supported by the Deutsche Forschungsgemeinschaft, the GeoForschungsZentrum Potsdam, and the Minerva Dead Sea Research Centre.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.R. (rumpker@gfz-potsdam.de).

Desert Seismology Group

M. Weber^{1,2}, K. Abu-Ayyash³, Z. Ben-Avraham⁴, R. El-Kelani⁵, Z. Garfunkel⁶, C. Haberland¹, A. Hofstetter⁷, R. Kind¹, J. Mechie¹, A. Mohsen¹, I. Qabbani², K. Wylegalla¹

¹GeoForschungsZentrum, Telegrafenberg, 14473 Potsdam, Germany; ²University of Potsdam, Potsdam, Germany; ³Natural Resources Authority, Amman, Jordan; ⁴Tel Aviv University, Tel Aviv, Israel; ⁵Ah-Najah National University, Nablus, Palestine Authority; ⁶Hebrew University, Jerusalem, Israel; ⁷Geophysical Institute of Israel, Lod, Israel

The oldest articulated chondrichthyan from the Early Devonian period

Randall F. Miller¹, Richard Cloutier² & Susan Turner^{1,3}

¹Steinhammer Palaeontology Laboratory, New Brunswick Museum, Saint John, New Brunswick E2K 1E5, Canada

²Laboratoire de Biologie évolutive, Université du Québec à Rimouski, Rimouski, Quebec G5L 3A1, Canada

³School of Geosciences, Monash University, Clayton, Victoria 3088, and Queensland Museum, South Brisbane, Queensland 4101, Australia

Chondrichthyans (including living sharks, skates, rays and chimaeras) have a fossil record of scales and dermal denticles perhaps dating back to the Late Ordovician period, about 455 million years ago^{1,2}. Their fossil tooth record extends to the earliest Devonian period, almost 418 million years ago³, whereas the oldest known articulated shark remains date from the Early Devonian period⁴, about 394 million years ago⁵. Here we report the discovery of an articulated shark that is almost 409 million years old⁵ from the Early Devonian (early Emsian) period of New Brunswick, Canada. The specimen, identified as *Doliodus problematicus* (Woodward)⁶, sheds light on the earliest chondrichthyans and their interrelationships with basal jawed

vertebrates. This species has been truly problematic⁷. Previously known only from isolated teeth^{2,6,8}, it has been identified as an acanthodian and a chondrichthyan. This specimen is the oldest shark showing the tooth families *in situ*, and preserves one of the oldest chondrichthyan braincases. More notably, it shows the presence of paired pectoral fin-spines, previously unknown in cartilaginous fishes.

Isolated and articulated Early to Middle Devonian shark specimens are rare^{1,9}. Until now, the oldest partial articulated shark, consisting of the braincase articulated with parts of the visceral skeleton, was *Pucapampella* from the Early Devonian of South Africa⁴. Significant Middle Devonian partially articulated specimens include *Pucapampella* from Bolivia^{10,11}, *Antarctilamna prisca* from Antarctica and Australia^{7,12}, and *Gladbachus adentatus* from Germany¹³.

Specimen NBMG (New Brunswick Museum, Geology) 10127/1a,b-4 consists of the anterior part of *D. problematicus*, forward of the mid-trunk region (Fig. 1). It is preserved dorsoventrally, oriented dorsal side up with exo- and endoskeletal elements preserved, including characteristic prismatic calcified cartilage, teeth, scales and large fin-spines. The specimen is cleaved in five parts, providing a series of transverse sections through the head and branchial region. The preserved length is 23 cm, suggesting a body length of perhaps 50–75 cm on the basis of shark comparative anatomy.

Prismatic calcified cartilage, considered to be a chondrichthyan synapomorphy^{1,14}, compose the neurocranium and splanchnocranium. The articulated jaws confirm that *D. problematicus* possessed tooth families and provide early evidence in chondrichthyans of the relationship of teeth to the dental lamina^{1,15}. Most teeth are partially buried; however, tooth families that are visible have teeth stacked in a row, with newer teeth sitting in a space representing the position of the dental lamina groove. Tooth bases abut a prominent dark-brown concave surface, interpreted as preserved basal connective tissue. The dentition shows weak dignathic and disjunct monognathic heterodonty, suggesting revision of earlier opinions about the evolution of shark teeth¹⁶. Functional upper and lower teeth, offset anteriorly, oppose one another with sharp lateral edges of principal cusps connecting in a scissors movement. The functional teeth show the asymmetry and range of variation previously recognized^{2,6,8}, and verify the position and number of tooth types in the jaw. Teeth are not seen in the symphyseal and parasymphyseal portions of the lower jaw.

The right side of the lower jaw shows about 15 tooth families; the left side has only 11 tooth families preserved, with bases of at least three anterior rows present in the cartilage. Tooth families expose up to three teeth each. Near the posterior jaw articulation, flat basal pads might represent the most posterior teeth. Lower tooth families are seen in cross-section, showing the apparently highly vascularized lower edge and new tooth germs. The last three to four posterior tooth families do not show dental membranes and thus are more like modified dermal scales. In a few teeth, two large divergent lateral outer cusps with two to four intermediate small cusps can be seen in cross-section. These and a thin section of a *D. problematicus* tooth from the National Museums of Scotland (RSM1897.51.46) show that the cusps are formed of orthodontine². Bases are rounded and cap-like with a row of five to six large foramina in the slightly concave foot. Cross-sections show osteodontine with a basal lamellar tissue, which directly abuts the dental membrane. The difference between the structure of the smaller posterior teeth (equivalent to type specimen BMNH (British Museum, Natural History) P.6540) and that of branchial denticles is still strong, contrary to one hypothesis on the origin of teeth¹⁷.

Woodward⁶ diagnosed the taxon “*Diplodus*” *problematicus* on an isolated tooth (BMNH P.6540), concluding that the diplodont (xenacanth) tooth type was present by Early Devonian. Traquair⁸

reassigned holotype and topotype specimens to *Doliodus*, with no doubt of its selachian nature, noting the succession of teeth. Since the early work and until recently^{2,7}, however, the teeth have been erroneously assigned to the Acanthodii¹⁸. *D. problematicus* teeth are now reassigned to Chondrichthyes, with a structure similar to that of some xenacanthiforms, and included in the Omalodontida Turner 1997, known from shark teeth with a labially extended base².

Until now, *Pucapampella* from the Late Emsian of South Africa⁴ and Emsian?/Late Eifelian/Givetian of Bolivia¹⁰ showed the oldest chondrichthyan cranium. The large braincase of *Doliodus*, in part and counterpart, is preserved from the precerebral fontanelle to just in front of the occipital region with the basal surface abutting the palatoquadrate (Fig. 2a, b). The precerebral fontanelle is prominent and large, similar to that of *Tamiobatis*⁹ and *Gladbachus*¹³. We considered its presence as a putative elasmobranch synapomorphy¹⁰ (neoselachian synapomorphy sensu Coates and Sequeira¹⁹); non-chondrichthyan taxa and holocephalans do not have a precerebral fontanelle^{10,19}. The neurocranium of *Doliodus* is the oldest to possess this fontanelle. The *Doliodus* neurocranium has a moderately long otico-occipital region and shows similar proportions to *Tamiobatis* and *Xenacanthus*⁹. Postorbital processes are wider than in *Pucapampella*, although narrower than the total width of the otico-occipital unit, a purported plesiomorphic condition

within chondrichthyans¹⁹. Poorly developed lateral otic processes are present, a character considered to be a chondrichthyan synapomorphy¹⁰. Evidence of a ventral otic fissure cannot be observed as preserved. Comparison of this feature to *Pucapampella*^{4,10} is not possible. Posteriorly, an elongate median endolymphatic fossa is present; this character sets *Doliodus* as a chondrichthyan above *Pucapampella*¹⁰.

The mandibular, hyoid and branchial arches are preserved almost in life-position (Fig. 1), but the latter two are slightly displaced posteriorly. The right side is more difficult to interpret because of the overpositioning and lateral compression of the meckelian cartilage, basihyal, palatoquadrate, ceratohyal and anterior three ceratobranchials. The inverted U-shaped median basihyal is wide and constricted at the symphysis. The basihyal and basibranchials are separated by a gap. Two basibranchials are preserved; the posterior one being the largest. Four (perhaps five) pairs of elongated, slightly sigmoid ceratobranchials form the main part of the branchial apparatus. The visceral skeleton of *Doliodus* shows gross similarity to that of *Gladbachus*¹³, with the exception of the shape of the basibranchials and ceratobranchial IV.

Morphology of both pectoral fins is well preserved (Figs 1 and 3). Uniquely for chondrichthyans, fin-spines form the anterior margin of the pectoral fins in *D. problematicus*. Notably, the fin-spine of the articulated shark *Antarctilamna prisca* (CPC (Commonwealth

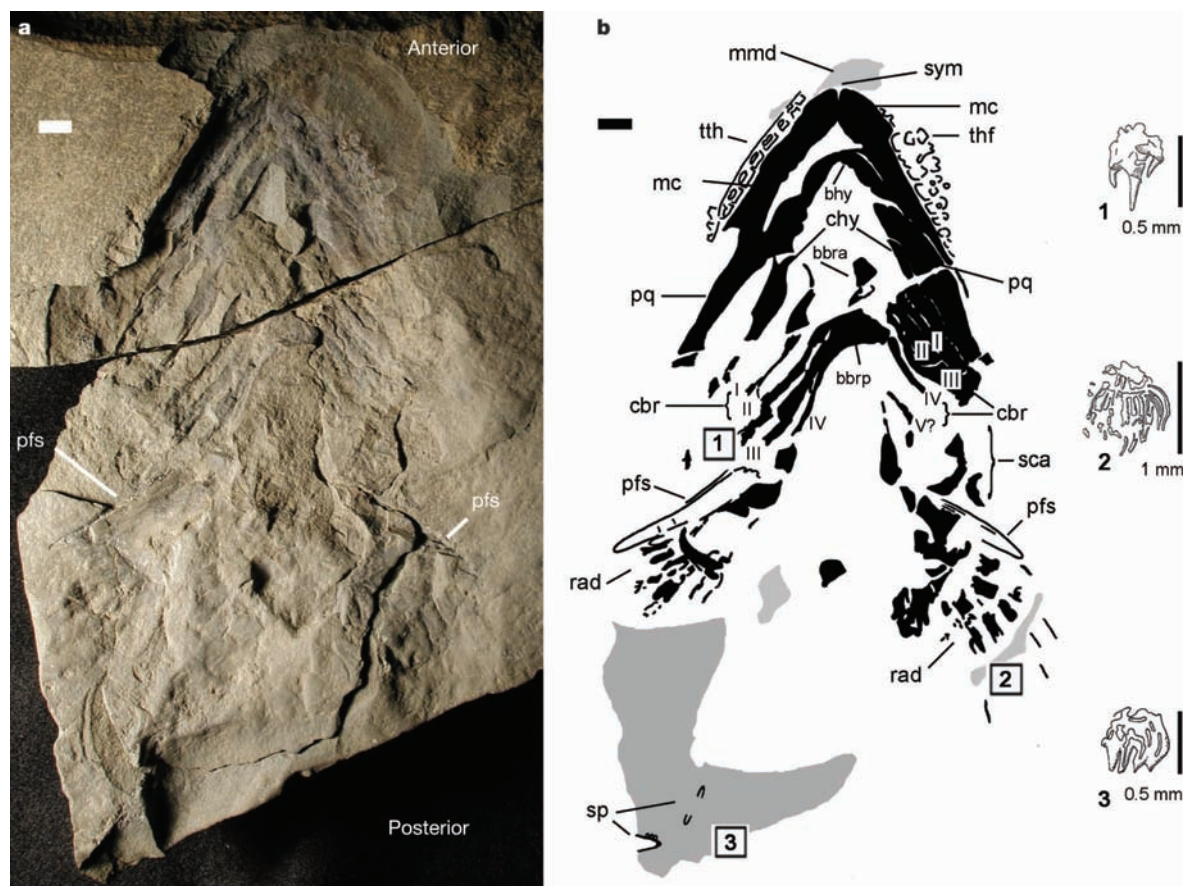


Figure 1 Partial articulated shark, *Doliodus problematicus* (NBMG 10127/1b,2,3). **a**, Specimen lying dorsal side up with the head at top and extending posteriorly to behind the pectoral fins. pfs, pectoral fin-spines. **b**, Map showing cartilage elements (black infill) and large areas of denticles (grey). mmd, location of mucous membrane denticles on counterpart; sym, symphysis; tth, area with *in situ* teeth; thf, *in situ* tooth family; mc, Meckel's cartilage; pq, palatoquadrate; bhy, basihyal; chy, ceratohyal;

bbra; anterior basibranchial; bbrp, posterior basibranchial; cbr, ceratobranchials (I–V?); sca, scapulocoracoid; pfs, pectoral fin-spines; rad, radials; sp, partial spines; denticle enlargements as preserved from branchial region (1), from pectoral fin (2) and from trunk region (3). Scale bar, 1 cm. Scales are separate for denticle enlargements.

Palaeontological Collection) 21187), which has been interpreted¹² as a dorsal spine, is suggestive of a pectoral fin-spine, being in the same position as the pectoral fin-spines described here. Until now, only dorsal fin-spines were known in basal chondrichthyans^{1,9}. Paired, dermal pectoral fin-spines were previously known only in placoderms, acanthodians and the basal osteichthyan *Psarolepis*²⁰. Their presence in *Doliodus* and perhaps *Antarctilamna* and the above-mentioned groups, suggests that it represents a gnathostome synapomorphy lost independently in Osteichthyes other than *Psarolepis*²¹, Placodermi and Chondrichthyes. The presence of large fin-spines associated with all fins except the caudal fin, or more specifically the presence of paired fin-spines, had been considered an acanthodian synapomorphy¹. This character can no longer be considered an acanthodian synapomorphy and the tenuous monophyly of acanthodians is now supported by a single scale histology synapomorphy¹.

The fin is aplesodic with radials extending half way to the margin (Figs 1 and 3). At least six (possibly seven) radials articulate on the lateral edge of a large basipterygial element (meso- or metapterygium); this feature is considered to be another chondrichthyan synapomorphy²². The area of dermal scales extends well lateral of the fin-spine distal tips, suggesting a large rounded pectoral fin. Left of the midline, near the posterior of the specimen, are small fin-spines that are furnished with hook-like denticles. These might be paired pelvic or intermediate spines, or parts of a collapsed dorsal fin-spine. Disarticulated fin-spines from the Campbellton Formation are identified as probable chondrichthyan and acanthodian^{6,8,23}. Pectoral fin-spines on NBMG 10127 are closest in size and ornament to those that were originally named "*Ctenacanthus*" *latispinosus*²³ and subsequently reassigned to *Climatius*⁶, a climatiid acanthodian¹⁸. Isolated *Climatius latispinosus* spines, NBMG 9986 and 10017, are preserved with *Doliodus* teeth and scales, and spine NMC (Canadian Museum of Nature) 12002 includes a nearby patch

of prismatic cartilage and branchial denticles. Further work is required to determine whether *D. problematicus* and *C. latispinosus* are synonymous.

Doliodus has rounded to polygonal, polyodontode mucous membrane denticles lining the inner upper palate and jaw edges (Fig. 2c) and rounded head denticles, multicuspid branchial denticles and ctenacanth-type complex trunk scales similar to those of *Antarctilamna* (Fig. 1), and thus is relatively advanced as compared with known purported Ordovician to Silurian sharks with simple placoid scales^{1,9}. Behind the branchial region shagreen extends posteriorly, with scattered denticles and dorsal scales infilling the central body area. Trunk scale morphology is most like scales described from the (Pragian) Jauf Formation in Saudi Arabia²⁴.

The Campbellton Formation²⁵, with its rich flora^{26,27} and terrestrial invertebrates²⁸, yields vertebrates in the lower "Atholville beds"²⁹. Miospores immediately below the articulated shark bed identify the *Emphanisporites annulatus*–*Camarozonotriletes sextantii* Assemblage Zone²⁶, corresponding to early, to early late Emsian (*dehiscens* to *serotinus* Conodont Zones) age. *Doliodus* teeth^{6,8} come from older beds at the base of the formation, perhaps near the 409-Myr-old Pragian/Emsian boundary⁵. The depositional environment has been considered fluvial²⁵; however, fossil assemblages^{26–28} suggest lagoonal and estuarine environments. Rare prasinophytes (tasmanids)²⁶ indicate a marine connection.

Discoveries of Early and Middle Devonian chondrichthyans from Gondwanan or neighbouring terranes have led to suggestions of a

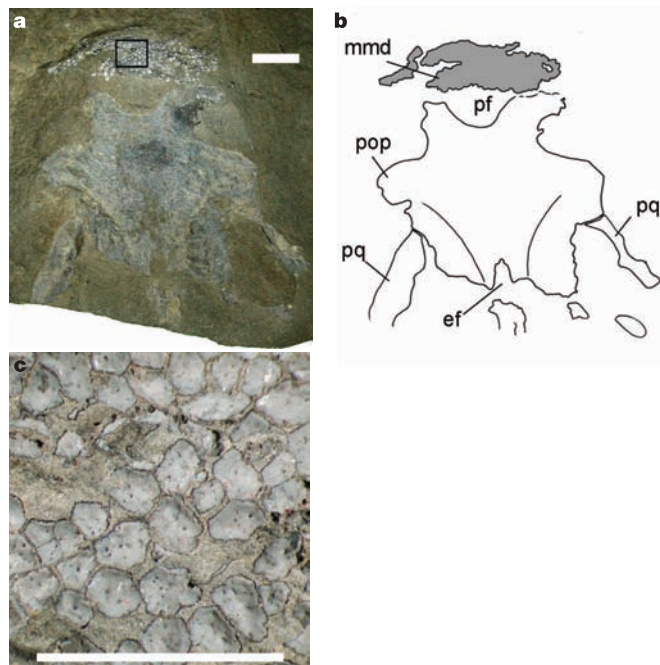


Figure 2 Neurocranium of *Doliodus problematicus* (NBMG 10127/4). **a**, Neurocranium and patch of mucous membrane denticles with enlarged area outlined. **b**, Map of neurocranium. pf, precerebral fontanelle; pop, postorbital process; ef, endolymphatic fossa; pq, palatoquadrate; mmd, mucous membrane denticles. Scale bar, 1 cm. **c**, Enlargement of mucous membrane denticles lining the inside of the mouth forward of the neurocranium. Scale bar, 0.5 cm.

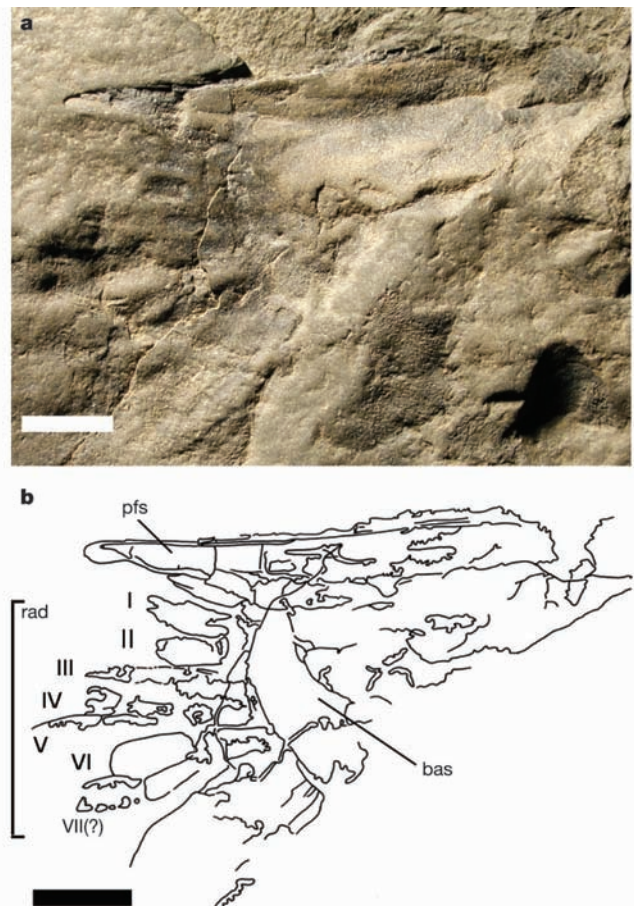


Figure 3 Left pectoral fin of *Doliodus problematicus* (NBMG 10127/3). **a**, Left pectoral fin. **b**, Map of left pectoral fin. pfs, pectoral fin-spine; rad, radials I–VII(?); bas, basipterygial element (meso- or metapterygium). Scale bar, 1 cm.

Gondwanan origin for sharks, although *D. problematicus* teeth presented a contradiction⁷. This specimen clearly places *D. problematicus* in Laurentia by the Early Devonian. Northern Gondwana and Laurentia were possibly close³⁰, across a shallow shelf connecting north Gondwanan shoreline locales. □

Received 12 February; accepted 11 August 2003; doi:10.1038/nature02001.

1. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, Oxford, 1996).
2. Turner, S. in *Recent Advances in the Origin and Early Radiation of Vertebrates* (eds Arratia, G., Wilson, M. V. H. & Cloutier, R.) (Verlag Dr F. Pfeil, Muenchen, in the press).
3. Mader, H. Schuppen und Zähne von Acanthodien und Elasmobranchiern aus dem Unterdevon Spaniens (Pisces). *Göttinger Arb. Geol. Paläont.* **28**, 1–58 (1986).
4. Maisey, J. G. & Anderson, M. E. A primitive chondrichthyan braincase from the Early Devonian of South Africa. *J. Vert. Paleontol.* **21**, 702–713 (2001).
5. Tucker, R. D. *et al.* New U-Pb zircon ages and the duration and division of Devonian time. *Earth Planet. Sci. Lett.* **158**, 175–186 (1998).
6. Woodward, A. S. On the Lower Devonian fish-fauna of Campbellton, New Brunswick. *Geol. Mag.* **9**, 1–6 (1892).
7. Long, J. A. & Young, G. C. Sharks from the Middle-Late Devonian Aztec Siltstone, southern Victoria Land, Antarctica. *Rec. West. Aust. Mus.* **17**, 287–308 (1995).
8. Traquair, R. H. Notes on the Devonian fishes of Campbelltown and Scaumenac Bay in Canada. *Geol. Mag.* **10**, 145–149 (1893).
9. Zangerl, R. in *Handbook of Paleichthyology* Vol. 3A *Chondrichthyes 1* (ed. Schultze, H.-P.) 1–114 (Gustav Fischer, Stuttgart, 1981).
10. Maisey, J. G. in *Major Events in Early Vertebrate Evolution. Palaeontology, Phylogeny, Genetics and Development*. Systematics Assoc. Spec. Vol. Ser. 61 (ed. Ahlberg, P. E.) 263–288 (Taylor and Francis, New York, 2001).
11. Janvier, P. & Suarez-Riglos, M. The Silurian and Devonian vertebrates of Bolivia. *Bull. Inst. Fr. Etud. Andines* **15**, 73–114 (1996).
12. Young, G. C. Devonian sharks from south-eastern Australia and Antarctica. *Palaeontology* **25**, 817–843 (1982).
13. Heidtke, U. H. J. & Krätschmer, K. *Gladbachus adentatus* nov. gen. et sp., ein primitiver Hai aus dem Oberen Givetium (Oberes Mitteldevon) der Bergisch Gladbach-Paffrath-Mulde (Rheinisches Schiefergebirge). *Mainzer geowissenschaftliche Mitteilungen* **30**, 105–122 (2001).
14. Maisey, J. G. Heads and tails: a chordate phylogeny. *Cladistics* **2**, 201–256 (1986).
15. Reif, W.-E. in *Evolutionary Biology* Vol. 15 (eds Hecht, M. K., Wallace, B. & Prance, G. T.) 287–368 (Plenum, New York, 1982).
16. Compagno, L. J. V. *Sharks of the Order Carcharhiniformes* (Princeton Univ. Press, Princeton, 1988).
17. Smith, M. M. & Coates, M. I. in *Major Events in Early Vertebrate Evolution. Palaeontology, Phylogeny, Genetics and Development*. Systematics Assoc. Spec. Vol. Ser. 61 (ed. Ahlberg, P. E.) 223–240 (Taylor and Francis, New York, 2001).
18. Denison, R. in *Handbook of Paleichthyology* Vol. 5 *Acanthodii* (ed. Schultze, H.-P.) 1–128 (Gustav Fischer, Stuttgart, 1979).
19. Coates, M. I. & Sequeira, S. E. K. The braincase of a primitive shark. *Trans. R. Soc. Edinb. (Earth Sci.)* **89**, 63–85 (1998).
20. Zhu, M., Yu, X. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397**, 607–610 (1999).
21. Zhu, M. & Schultze, H.-P. in *Major Events in Early Vertebrate Evolution. Palaeontology, Phylogeny, Genetics and Development*. Systematics Assoc. Spec. Vol. Ser. 61 (ed. Ahlberg, P. E.) 289–314 (Taylor and Francis, New York, 2001).
22. Coates, M. I. & Sequeira, S. E. K. in *Major Events in Early Vertebrate Evolution. Palaeontology, Phylogeny, Genetics and Development*. Systematics Assoc. Spec. Vol. Ser. 61 (ed. Ahlberg, P. E.) 241–262 (Taylor and Francis, New York, 2001).
23. Whiteaves, J. F. On some fossil fishes Crustacea & Mollusca from the Devonian rocks at Campbellton, NB, with descriptions of five new species. *Can. Nat.* **10**, 93–101 (1881).
24. Forey, P. L., Young, V. T. & McClure, H. A. Lower Devonian fishes from Saudi Arabia. *Bull. Br. Mus. Nat. Hist. (Geol.)* **48**, 25–43 (1992).
25. Williams, G. L., Fyffe, L. R., Wardle, R. J., Colman-Sadd, S. P. & Boegner, R. C. *Lexicon of Canadian Stratigraphy* Vol. VI *Atlantic Region* (Can. Soc. Petroleum Geologists, Calgary, 1985).
26. Blicke, A. & Cloutier, R. in *Palaeozoic Vertebrate Biochronology and Global Marine/Non-Marine Correlation* (eds Blicke, A., Turner, S.) *Cour. Forsch.-Inst. Senckenberg* **223**, 223–269 (2000).
27. Gensel, P. G. & Andrews, H. N. *Plant Life in the Devonian* (Praeger, New York, 1984).
28. Shear, W., Gensel, P. & Jeram, A. Fossils of large terrestrial arthropods from the Lower Devonian of Canada. *Nature* **384**, 555–557 (1996).
29. Dineley, D. L. & Williams, B. P. J. The Devonian continental rocks of the lower Restigouche River, Quebec. *Can. J. Earth Sci.* **5**, 945–953 (1968).
30. Keppie, J. D. & Ramos, V. A. in *Laurentia-Gondwana Connections Before Pangea* Geol. Soc. Am. Spec. Pap. 336 (eds Ramos, V. A. & Keppie, J. D.) 267–276 (Geol. Soc. Am., Colorado, 1999).

Acknowledgements We thank J. McGovern, S. Miller, H. Wilson, the other members of the field team who discovered the specimen and C. Burrow for discussion on acanthodian spines. This research was supported in part by the George Frederic Matthew Research Grant of the New Brunswick Museum (to R.C. and S.T.) and a grant from the Natural Sciences and Engineering Research Council (to R.C.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.E.M. (millerrf@nb.aibn.com).

Temperature excludes N₂-fixing heterocystous cyanobacteria in the tropical oceans

Marc Staal¹, Filip J. R. Meysman² & Lucas J. Stal¹

¹Department of Marine Microbiology, and ²Department of Ecosystem Studies, NIOO-KNAW, PO Box 140, 4400 AC Yerseke, The Netherlands

Whereas the non-heterocystous cyanobacteria *Trichodesmium* spp. are the dominant N₂-fixing organisms in the tropical oceans¹, heterocystous species dominate N₂ fixation in freshwater lakes and brackish environments such as the Baltic Sea². So far no satisfactory explanation for the absence of heterocystous cyanobacteria in the pelagic of the tropical oceans has been given, even though heterocysts would seem to represent an ideal strategy for protecting nitrogenase from being inactivated by O₂, thereby enabling cyanobacteria to fix N₂ and to perform photosynthesis simultaneously. *Trichodesmium* is capable of N₂ fixation, apparently without needing to differentiate heterocysts³. Here we show that differences in the temperature dependence of O₂ flux, respiration and N₂ fixation activity explain how *Trichodesmium* performs better than heterocystous species at higher temperatures. Our results also explain why *Trichodesmium* is not successful in temperate or cold seas. The absence of heterocystous cyanobacteria in the pelagic zone of temperate and cold seas, however, requires another explanation.

As primary production in vast areas of the oceans is predominantly controlled by the availability of nitrogen, biological N₂ fixation could overcome this limitation¹. But N₂ fixation in the marine pelagic environment seems to be mainly restricted to (sub) tropical regions. The organisms responsible for most of the N₂ fixation in the tropical oceans are *Trichodesmium* spp., filamentous non-heterocystous cyanobacteria that can form massive surface blooms^{4,5}. Although free-living heterocystous cyanobacteria are reported to be present in the marine pelagic environment, their numbers are low and presumably they show very low growth rates⁶. This is notable because heterocystous cyanobacteria are considered to be better adapted to diazotrophic growth than are non-heterocystous species^{1,7}, as heterocysts (differentiated cells enveloped by a glycolipid layer in which N₂ fixation takes place) are assumed to be effective in protecting the N₂-fixing enzyme nitrogenase from inactivation by O₂ (refs 8, 9).

The absence of heterocystous cyanobacteria in the marine pelagic environment contrasts strongly with their presence in freshwater lakes and brackish environments, where they can form dense blooms. Although heterocystous cyanobacteria can thrive in marine tropical systems, they are found mostly as epiphytes, in symbiosis with the planktonic diatom *Rhizosolenia* or in microbial mats¹⁰. But these specific environments are regularly oversaturated with O₂ during the daytime, and therefore provide conditions that are different from those experienced by free-living organisms in pelagic systems. The expected higher O₂ fluxes require a better protection of nitrogenase.

This leaves us with two main questions regarding the global distribution of N₂ fixation. First, why are free-living heterocystous cyanobacteria not the dominant N₂-fixing organisms in the tropical oceans? Second, why are *Trichodesmium* spp. not able to thrive in marine, brackish or even freshwater environments in temperate and polar regions? Here, we propose that a glycolipid cell envelope, which acts as an effective diffusion barrier for O₂ in heterocysts, does not provide an advantage in sea water at increased temperatures, and thus heterocystous cyanobacteria are out-competed by *Trichodesmium* spp. Our results also explain why *Trichodesmium*

spp., which possess specialized N_2 -fixing cells with high respiration rates but without a glycolipid layer (diazocytes)^{3,11}, are not successful in temperate or cold seas.

Measurements of nitrogenase activity by the acetylene reduction assay in the heterocystous *Nodularia spumigena* strain CCY 9414, *Anabaena* sp. strain CCY 9901 and *Trichodesmium* sp. strain IMS101 showed that nitrogenase activity in the dark (N_d) and in the light (N_{tot}) increased with temperature (Fig. 1). All temperature dependence (Q_{10}^*) values for N_d were close to 1.1 (Table 1), except for that of *Trichodesmium* sp., which still had a Q_{10}^* of 1.12 for N_d at 20–35 °C, but a Q_{10}^* of 2.06 for N_d at 15–20 °C. An even higher Q_{10}^* (4.7) was found at 10–15 °C for *N. spumigena* grown at 20 °C.

In the light, photosynthetic electron transport adds to the respiratory production of ATP; therefore, at saturating irradiances nitrogenase will no longer be limited by ATP and will achieve its maximum attainable activity (N_{tot})¹². The Q_{10}^* of N_{tot} was measured at 20% O_2 saturation for the same strains (Fig. 1). Table 1 shows that all values of Q_{10}^* were close to 2, a value found for many enzymatic processes^{13,14}. When N_{tot} does not increase notably under anaerobic conditions, it can be considered to be the potential nitrogenase activity¹⁴. Conducting the same temperature experiments under anaerobic conditions, we found that N_{tot} increased by 12 ± 3 , 29 ± 15 , 2 ± 4 and $12 \pm 9\%$ in *N. spumigena* grown at 20 °C, *N. spumigena* grown at 25 °C, *Anabaena* sp. and *Trichodesmium* sp., respectively. Thus, N_{tot} closely matched the potential nitrogenase activity, except for in *N. spumigena* grown at 25 °C. This last strain seemed to become moderately limited by reducing equivalents when measured at 25 °C and higher.

Owing to a limited influx of O_2 in N_2 -fixing cells, respiratory ATP production limits N_2 fixation in the dark^{9,12}. Because the flux of O_2 into heterocysts and diazocytes is crucial for regulating N_2 fixation, we considered the factors on which it depends. For this purpose, we developed a simplified reactive-transport model (for simplicity, we modelled both heterocysts and diazocytes as spherical cells; the model shows that the cylindrical shape of groups of diazocyte cells does not essentially alter our analysis; see Supplementary Appendices A and B).

For a heterocyst, we assumed that main resistance to diffusion lies in the glycolipid layer, and thus the oxygen flux J can be approxi-

mated (see Supplementary Appendix A) by

$$J = 4\pi DC_\infty \left\{ \epsilon R \left(1 + \frac{R}{L_g} \right) \right\} \quad (1)$$

where R is the cell radius, L_g is the thickness of the glycolipid layer, C_∞ is the external concentration of dissolved O_2 , and D is the gas diffusion coefficient. The efficiency of the glycolipid layer as a gas diffusion barrier (ϵ) is defined as the ratio of the diffusion coefficient in the glycolipid layer to the diffusion coefficient in water.

By contrast, a diazocyte does not possess a glycolipid layer (that is, $L_g = 0$), and accordingly there is no resistance to diffusion in the cell wall (that is, $\epsilon = 1$). The oxygen flux J equation (1) thus reduces (see Supplementary Appendix B) to

$$J = 4\pi DC_\infty \left\{ R \left(1 + \frac{R}{L_b} \right) \right\} \quad (2)$$

where L_b is the thickness of the boundary layer.

Equations (1) and (2) can now be used to assess the influence of environmental conditions. In a first-order approach, it seems reasonable to consider that the geometrical parameters R , L_g and L_b are independent of temperature and salinity; however, the external concentration of dissolved O_2 (C_∞) and the gas diffusion coefficient (D) are both strongly dependent on these parameters (see http://www.unisense.com/support/pdf/gas_tables.pdf). In addition, we neglect any changes in the permeability of the cell membranes due to temperature or salinity. Under these assumptions, the Q_{10}^* relation for the oxygen flux eventually becomes the same for both heterocysts and diazocytes, and is given by

$$Q_{10}^*(J) = Q_{10}^*(D) \times Q_{10}^*(C_\infty) \quad (3)$$

Applying equation (3) shows that, despite the lower concentration of dissolved O_2 at increasing temperature ($Q_{10}^*(D) \approx 0.84$), the O_2 flux increases with an overall $Q_{10}^*(J)$ value of 1.08. This value closely approaches the Q_{10}^* values obtained for N_d (Table 1). Our model shows that N_2 fixation in the dark in heterocysts and diazocytes is limited by respiration, which is eventually controlled by O_2 diffusion. It also indicates that our assumption that the permeability of the glycolipid layer does not change with temperature is valid.

We now can distinguish two types of control of N_d : ‘transport control’, when N_d is limited by the flux of O_2 into the cell; and ‘reaction control’, when N_d is limited by the activity of respiratory enzymes or potential nitrogenase activity. Both controls have a markedly different temperature response. As discussed above, the Q_{10}^* of the O_2 flux is moderate (~ 1.1) and determined by physical properties (http://www.unisense.com/support/pdf/gas_tables.pdf). By contrast, the Q_{10}^* of the respiratory activity is reported to be more pronounced (~ 2 ; ref. 14). Consequently, we advance a conceptual scheme in which the temperature response curve of N_d is composed of two phases (Fig. 2).

In the first phase, the enzyme activity is limiting (reaction control), whereas in the second phase the O_2 supply is limiting (transport control). Note that when respirational activity becomes

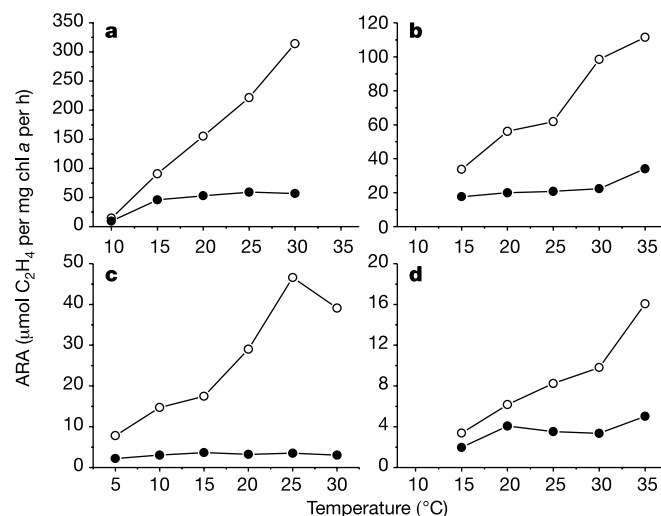


Figure 1 Relationship of nitrogenase activity versus temperature measured at 20% O_2 . Nitrogenase activity (acetylene reduction, ARA) was measured in the dark (N_d , filled circles) and at saturating irradiances (N_{tot} , open circles) for *N. spumigena* (strain CCY 9414) cultured at 20 °C (a) and 25 °C (b), *Anabaena* sp. (strain CCY 9901) cultured at 20 °C (c) and *Trichodesmium* sp. (strain IMS101) cultured at 25 °C (d).

	Temperature range	$Q_{10}^*(N_d)$	$Q_{10}^*(N_{tot})$
<i>N. spumigena</i> (20 °C) (strain CCY 9414)	10–15 °C	4.72	1.96
	15–30 °C	1.16	2.26
<i>N. spumigena</i> (25 °C) (strain CCY 9414)	15–35 °C	1.16	1.81
<i>Anabaena</i> sp. (strain CCY 9901)	5–30 °C	1.08	2.34
<i>Trichodesmium</i> sp. (strain IMS101)	15–20 °C	2.06	1.64
	20–35 °C	1.12	1.84

Q_{10}^* values for N_d and N_{tot} were measured for *N. spumigena* (strain CCY 9414) cultured at 20 °C and 25 °C, *Anabaena* sp. (strain CCY 9901) cultured at 20 °C, and *Trichodesmium* sp. (strain IMS101) cultured at 25 °C.

limiting, anoxia can be no longer maintained and nitrogenase will be irreversibly inactivated, prohibiting N_2 fixation. Under transport control, a further rise in temperature will widen the gap between effective respiration (controlled by the O_2 flux) and potential respiration (governed by the respiratory enzyme system). This has been effectively shown for cultures of *N. spumigena* and *Anabaena* sp., in which the optimum O_2 concentration for N_2 fixation increased at higher temperature (see Supplementary Appendix C). The O_2 optimum for N_2 fixation in *Trichodesmium* sp. (strain IMS 101) was much lower with about 10% O_2 at 25 °C.

Linear fitting of N_{tot} and N_d in the transport-controlled temperature range showed that the intercepts of these lines occurred at 16 °C, 11–12 °C and –7 °C for *Trichodesmium* sp., *N. spumigena* and *Anabaena* sp., respectively. Because N_{tot}/N_d is greater than 1 at these intercepts, we can conclude that the respiratory system becomes reaction controlled at a higher temperature than does nitrogenase. The intercept coincides with the distribution of *Trichodesmium* sp. in the ocean¹ and explains the difference in Q_{10}^* between the 15–20 °C and 20–35 °C temperature intervals. The same holds for *N. spumigena*, which proliferates in the Baltic Sea when the water temperature rises above 10–15 °C. *Anabaena* sp. is never a dominant species in the Baltic Sea, but notably this species appears early in the year when the water temperature is still low².

The potential nitrogenase activity (N_{tot}) increased relatively more with temperature than did N_d , resulting in an increase in the ratio N_{tot}/N_d (Fig. 3) with temperature. The ratio N_{tot}/N_d (≥ 1) can be used as an indicator of the relative ATP limitation of nitrogenase activity in the dark¹². A higher N_{tot}/N_d indicates a higher ATP limitation of nitrogenase in the dark. When *N. spumigena* was grown at 25 °C and nitrogenase activity was measured at the same temperature, N_{tot}/N_d was 2.5. A similar value (2.9) was found for *N. spumigena* grown and assayed at 20 °C, whereas a value of 3.8 was found for cultures grown at 20 °C and measured at 25 °C (Fig. 3). It seems that this species is able to adapt. In experiments with *N. spumigena* other than those described here, an N_{tot}/N_d of 2.57 ± 0.69 ($n = 18$) was found.

We examined 21 different heterocystous species isolated from various marine, brackish and freshwater environments (for strains, see Supplementary Appendix D). The average N_{tot}/N_d at 20 °C for all of these organisms was 3.20 ± 1.07 . Thus, it seems that there is an optimal N_{tot}/N_d in heterocystous cyanobacteria in the range of 2.5–4. Given this optimal value and the shift in N_{tot}/N_d for *N. spumigena* in response to a long-term change in temperature,

we can speculate on how heterocystous cyanobacteria can adjust N_{tot}/N_d . One option is to decrease N_{tot} , either by lowering the concentration of nitrogenase in the heterocyst or by decreasing the supply of reducing equivalents. Both options would decrease the fixation rate of N_2 and thus reduce the efficiency of the heterocyst, thereby impairing the organism's success.

A more likely option is to increase N_d . Assuming transport control, this can be achieved either by raising C_{∞} , D or ε , or by decreasing the glycolipid layer thickness (L_g). As C_{∞} and D are physical properties beyond the control of the organism, the only physiological way to change N_{tot}/N_d is to improve cell wall permeability. This will induce a higher influx of O_2 , resulting in higher respiration rates and ATP production, and consequently in an increase in N_d . Changes in cell wall thickness and permeability have been observed in heterocystous cyanobacteria growing under different O_2 concentrations^{15,16}. If the increase in permeability of the heterocyst envelope were the strategy to respond to an increase in temperature, it would mean that heterocysts would eventually lose their selective advantages of formation and maintenance over non-heterocystous cyanobacteria such as *Trichodesmium* sp., which do not devote energy and resources to such a gas diffusion barrier.

Using equation (1), we calculated that the flux of O_2 is increased by about 25% when the salinity is decreased from 35 (sea water) to 0 (fresh water). This emphasizes the need for an efficient diffusion barrier for the N_2 -fixing cells in fresh or brackish waters. The same is true for marine environments, which are frequently oversaturated with O_2 . This explains the presence of heterocystous cyanobacteria in these environments.

The measurement of dark N_2 fixation allows us to estimate the relative gas diffusion into the N_2 -fixing cell. Using equation (1) of the heterocyst reactive-transport model, we calculated an N_2 flux that is 1.75 times the O_2 flux when the N_2 concentration inside the heterocyst is set at zero. The actual intracellular concentration of N_2 will be higher, however, resulting in a lower N_2 flux. The actual N_2 and O_2 fluxes in the heterocyst are therefore probably almost similar. Calculating the Q_{10}^* of the N_2 flux, we found a value similar to that of the O_2 flux (1.09). From this we conclude that the advantage to the cyanobacterium gained by the ability of the heterocyst cell wall to decrease the flux of O_2 is counteracted by a simultaneous limitation of the N_2 flux. Thus, the gas diffusion barrier must be a trade-off between the N_2 required by nitrogenase in the light (N_{tot}) and the maximum capacity to respire O_2 to maintain anoxia. This explains why heterocystous cyanobacteria show only a relatively small range around an optimal value of N_{tot}/N_d . The N_{tot}/N_d found for *Trichodesmium* sp. was 2.33, which is close to the 'optimum' ratio found for heterocystous cyanobacteria.

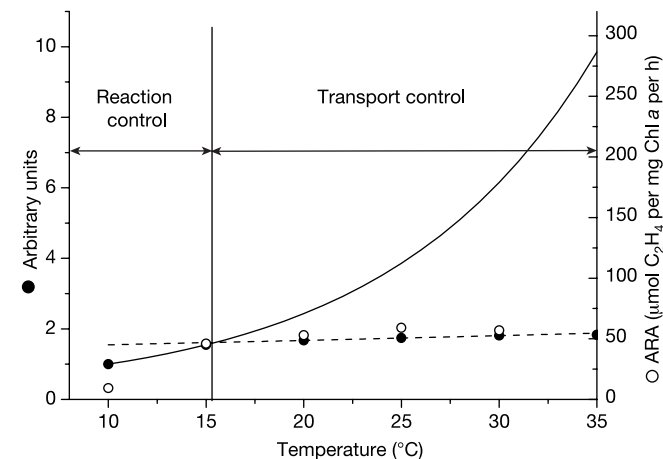


Figure 2 Theoretical enzyme activity curves with a Q_{10}^* of 2 (unbroken line) and a Q_{10}^* of 1.16 (broken line). Filled circles indicate the theoretical activity curve, which would be measured according to theory; open circles show the measured N_d value in *N. spumigena* grown at 20 °C.

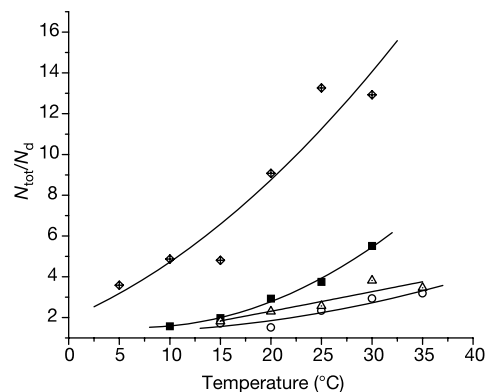


Figure 3 Changes in the ratio of N_{tot} to N_d with temperature. Nitrogenase activity (acetylene reduction) was measured at 20% O_2 for *N. spumigena* (strain CCY 9414) cultured at 20 °C (squares) and 25 °C (triangles), *Anabaena* sp. (strain CCY 9901) cultured at 20 °C (diamonds) and *Trichodesmium* sp. (strain IMS101) cultured at 25 °C (circles).

Our results support a simple satisfactory explanation for the presence of the non-heterocystous cyanobacteria *Trichodesmium* spp. in the (sub) tropical oceans and their absence in temperate and cold seas as well as in freshwater and brackish environments. Although we have also provided an explanation for the exclusion of free-living heterocystous cyanobacteria in the nitrogen-depleted euphotic zone of the pelagic tropical ocean, our findings do not explain why these organisms are virtually absent from temperate and polar marine waters. It seems that other factors prevent the proliferation of N₂-fixing cyanobacteria in these environments. We can speculate about possible candidates such as the availability of iron or phosphorus, but a sound explanation cannot be offered at present. □

Methods

Strains

Nodularia spumigena (strain CCY 9414) and *Anabaena* sp. (strain CCY 9901) were grown at 20 °C in medium free of combined nitrogen, comprising one part ASN3 and two parts BG 11 (ref. 17), giving a salinity of 11. *Nodularia spumigena* was also grown at 25 °C. For comparison, we measured nitrogenase activity in 19 other heterocystous cyanobacteria (see Supplementary Information). *Trichodesmium* sp. (IMS101) was grown at 25 °C in YBC II medium¹⁸.

Nitrogenase activity and irradiance curves

Nitrogenase activity was measured by an on-line acetylene reduction assay equipped with a gas chromatograph (GC-FID)¹⁹. Nitrogenase versus irradiance curves were determined with the cyanobacteria immobilized on glass fibre filters (Whatman) in a temperature-controlled incubator¹⁹. For each temperature at which a measurement was made, nitrogenase versus irradiance curves were recorded at 20% O₂. A slide projector with a set of ten neutral density filters (Balzers) provided photon irradiances from 0 to 1,600 μmol m⁻² s⁻¹. Light response curves were fitted in Origin 6.0 (Microcal) by nonlinear least of squares fitting, using the Levenberg–Marquardt algorithm with the rectangular hyperbola model^{12,20}

$$N_m \left(\frac{\alpha I}{N_m + \alpha I} \right) + N_d$$

where N_m is the nitrogenase activity at saturating irradiances minus N_d , α is the light affinity coefficient for nitrogenase activity, and I is the photon irradiance. N_{tot} was calculated by summing N_d and N_m (ref. 12).

Other parameters

The O₂ optima of acetylene reduction assay at the different incubation temperatures were determined by measuring N₂ fixation at different O₂ concentrations (0, 2.5, 5, 10, 20 and 30% O₂) at each photon irradiance. The concentration of O₂ and the diffusion coefficients used in equations (1) and (2) were taken from the tables of Seawater and Gases (http://www.unisense.com/support/pdf/gas_tables.pdf). The model parameters for the boundary layer thickness and cell diameter were arbitrary values. We tested the robustness of the model towards these parameters. Changing their values did not alter the relative fluxes at the different salinities and temperatures.

We calculated Q_{10}^* by fitting the temperature response curves with the equation $Y = Y_0 \exp(\lambda T)$, where $Q_{10}^* = \exp(\lambda 10)$. The Q_{10}^* for *Trichodesmium* IMS101 at 15–20 °C and for *N. spumigena* at 10–15 °C was estimated with a linear function, because an exponential fit through two points may be a source of erroneous results.

Chlorophyll *a* was extracted with 96% ethanol for 24 h in the dark. Absorption was read at 665 nm with a spectrophotometer, and the chlorophyll concentration was calculated by using an absorption coefficient of 72.3 ml mg⁻¹ cm⁻¹.

Received 30 April; accepted 11 August 2003; doi:10.1038/nature01999.

- Karl, D. *et al.* Dinitrogen fixation in the world's oceans. *Biogeochemistry* **57/58**, 47–98 (2002).
- Laamanen, M. Environmental factors affecting the occurrence of different morphological forms of cyanoprokaryotes in the northern Baltic Sea. *J. Plankt. Res.* **19**, 1385–1403 (1997).
- Berman-Frank, I. *et al.* Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* **294**, 1534–1537 (2001).
- Berman-Frank, I., Cullen, J. T., Shaked, Y., Sherrill, R. M. & Falkowski, P. G. Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. *Limnol. Oceanogr.* **46**, 1249–1260 (2001).
- Sáudo-Wilhelmy, S. A. *et al.* Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* **411**, 66–69 (2001).
- Carpenter, E. J. & Janson, S. *Anabaena gerdii* sp. nov., a new planktonic filamentous cyanobacterium from the South Pacific Ocean and Arabian Sea. *Phycologia* **40**, 105–110 (2001).
- Bergman, B., Gallon, J. R., Rai, A. N. & Stal, L. J. N₂ fixation by non-heterocystous cyanobacteria. *FEMS Microb. Rev.* **19**, 139–185 (1997).
- Walsby, A. E. The permeability of heterocysts to the gases nitrogen and oxygen. *Proc. R. Soc. Lond. B* **226**, 345–366 (1985).
- Wolk, C. P., Ernst, A. & Elhai, J. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 770–823 (Kluwer Academic, Dordrecht, 1994).
- Hoffmann, L. Marine cyanobacteria in tropical regions: diversity and ecology. *Eur. J. Phycol.* **34**, 371–379 (1999).
- Kana, T. M. Rapid oxygen cycling in *Trichodesmium thieboutii*. *Limnol. Oceanogr.* **38**, 18–24 (1993).
- Staal, M., te Lintel Hekkert, S., Herman, P. & Stal, L. J. Comparison of models describing light

dependency of N₂ fixation in heterocystous cyanobacteria. *Appl. Environ. Microbiol.* **68**, 4679–4683 (2002).

- Stal, L. J. & Walsby, A. E. Photosynthesis and nitrogen fixation in a cyanobacterial bloom in the Baltic Sea. *Eur. J. Phycol.* **35**, 97–108 (2000).
- Davison, I. R. Environmental effects on algal photosynthesis: temperature. *J. Phycol.* **27**, 2–8 (1991).
- Kangatharalingam, N., Dodds, W. K., Priscu, J. C. & Paele, H. W. Nitrogenase activity, photosynthesis, and the degree of heterocyst aggregation in the cyanobacterium *Anabaena flos-aquae*. *J. Phycol.* **27**, 680–686 (1991).
- Murry, M. A. & Wolk, C. P. Evidence that the barrier to the penetration of oxygen to heterocysts depends upon two layers of the cell envelope. *Arch. Microbiol.* **151**, 469–474 (1989).
- Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **111**, 1–61 (1979).
- Chen, Y. B., Zehr, J. P. & Mellon, M. Growth and nitrogen fixation of the diazotrophic filamentous non-heterocystous cyanobacterium *Trichodesmium* sp. IMS 101 in defined media: evidence for a circadian rhythm. *J. Phycol.* **32**, 916–923 (1996).
- Staal, M., te Lintel Hekkert, S., Harren, F. & Stal, L. J. Nitrogenase activity in cyanobacteria measured by the acetylene reduction assay: a comparison between batch incubation and on-line monitoring. *Environ. Microbiol.* **3**, 343–351 (2001).
- Talling, J. F. Photosynthetic characteristics of some freshwater plankton diatoms in relation to underwater radiation. *New Phytol.* **56**, 29–50 (1957).

Supplementary Information accompanies the paper on www.nature.com/nature.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (M.Staal@nioo.knaw.nl).

Jelly belly protein activates the receptor tyrosine kinase Alk to specify visceral muscle pioneers

Hsiu-Hsiang Lee¹, Audra Norris², Joseph B. Weiss^{2*} & Manfred Frasch^{1*}

¹Brookdale Department of Molecular, Cell and Developmental Biology, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1020, New York, New York 10029, USA

²Molecular Medicine and Cardiology, Oregon Health and Sciences University, 3181 SW Sam Jackson Park Road/NRC3, Portland, Oregon 97201-1807, USA

*These authors contributed equally to this work

The secreted protein Jelly belly (Jeb) is required for an essential signalling event in *Drosophila* muscle development. In the absence of functional Jeb, visceral muscle precursors are normally specified but fail to migrate and differentiate¹. The structure and distribution of Jeb protein implies that Jeb functions as a signal to organize the development of visceral muscles¹. Here we show that the Jeb receptor is the *Drosophila* homologue of anaplastic lymphoma kinase (Alk), a receptor tyrosine kinase of the insulin receptor superfamily. Human ALK was originally identified as a proto-oncogene, but its normal function in mammals is not known². In *Drosophila*, localized Jeb activates Alk and the downstream Ras/mitogen-activated protein kinase cascade to specify a select group of visceral muscle precursors as muscle-patterning pioneers. Jeb/Alk signalling induces the myoblast fusion gene *dumbfounded* (*duf*; also known as *kirre*) as well as *org-1*, a *Drosophila* homologue of mammalian *TBX1*, in these cells.

Signalling molecules and their receptors orchestrate cell fate decisions essential to organogenesis. Studies of mesoderm development in *Drosophila* have highlighted the role of evolutionarily conserved signalling systems, and the transcription factors they regulate, in the elaboration of the mesoderm into its derivative tissues. The earliest cell fate assignments in the mesoderm are coordinated by inductive signals from the ectoderm. Decapentaplegic (Dpp), a *Drosophila* BMP signal, induces subjacent dorsal

mesoderm to express Tinman (Tin), a homeodomain protein essential for heart, visceral and dorsal somatic mesoderm development^{3–5}. Dpp and Tin, together with Hedgehog, induce visceral mesoderm by activating the expression of two transcription factors, Bagpipe (Bap) and Biniou (Bin)^{5–8}. A third signal, Wingless, antagonizes these visceral mesoderm-inducing activities^{7,9,10}. The combined actions of ectodermally derived Dpp, Hedgehog and Wingless generate segmental clusters of visceral mesoderm precursors in the dorsal mesoderm.

The secreted protein Jeb is necessary for the subsequent rearrangement of these segmental clusters of visceral mesoderm precursors into bilateral longitudinal bands and for visceral muscle differentiation¹. Jeb is produced in ventral somatic mesoderm, locally secreted, and is specifically taken up by the visceral mesoderm cells¹. Its detailed developmental role, however, has not been defined. We show that one critical function of Jeb signalling is to subdivide the pool of visceral mesoderm precursors into two distinct subtypes: muscle founders and fusion-competent cells (Fig. 1). This subdivision is key to the muscle specification and fusion pathway, a hierarchical system for patterning muscles (reviewed in refs 11, 12). As first shown for somatic muscle development in *Drosophila*, founder myoblasts are patterning pioneers^{13,14}. They establish specific muscles and recruit fusion-competent myoblasts to fuse with them into mature syncytial muscle fibres. Founder myoblasts and fusion-competent myoblasts are identified by the expression of functional components of the myoblast fusion pathway. Founder cells express Duf, a transmembrane protein necessary for recruitment of fusion-competent cells. Fusion-competent cells express Sticks and stones (Sns), a transmembrane protein also required for fusion.

Besides *duf*^{15,16}, we have found that the expression of a *Drosophila* homologue of mammalian TBX1—*org-1* (ref. 17), initially detectable in all visceral mesoderm precursors (Fig. 1a)—is rapidly restricted to founder cells (Fig. 1b, c). TBX1, a T-box transcription

factor, is required for cardiovascular development, and in humans haploinsufficiency of its function results in the congenital cardiovascular abnormalities associated with 22q deletions in DiGeorge/velocardiofacial syndrome (reviewed in ref. 18). As shown in Fig. 1, the expression of both *org-1* and *duf* markedly responds to Jeb. Both are absent from the visceral mesoderm in stage 11–12 *jeb* mutant embryos (Fig. 1e, h), whereas ectopic expression of Jeb activates them throughout the visceral mesoderm (Fig. 1f, i). A third gene, *sns*, normally expressed in fusion-competent cells in a pattern reciprocal to *duf* is negatively regulated by Jeb (see Supplementary Fig. 1a–c).

Positive regulation of *duf* and negative regulation of *sns* implies that Jeb signalling specifies visceral mesoderm founders. As assayed by the markers *duf*, *org-1* and *sns*, no visceral muscle founders are specified in *jeb* mutant embryos. Instead all visceral mesoderm precursors become fusion-competent myoblasts. The consequence of absent visceral mesoderm founders, as shown by cell-lineage experiments (Fig. 1j, k; see also Supplementary Fig. 1d, e), is fusion of visceral fusion-competent myoblasts with somatic muscle founders and loss of visceral musculature. Somatic muscle patterning, however, is unaffected¹.

Localized activation of the Ras/mitogen-activated protein kinase (MAPK) cascade in the visceral mesoderm has been noted previously¹⁹. In the somatic muscle lineage this pathway is required for founder cell specification²⁰. We therefore hypothesized that Jeb signals through the Ras/MAPK cascade in the visceral mesoderm. As shown in Figs 2a and 3a, b, activated MAPK is indeed detected in the visceral mesoderm precursors that take up Jeb. The observed overlapping signals for diphospho-MAPK and *org-1*, as well as the exclusive staining patterns for diphospho-MAPK and *sns* (Fig. 2b and data not shown), confirm that the MAPK pathway is activated in presumptive visceral muscle founders. Moreover, Jeb signalling is necessary and sufficient to activate the Ras/MAPK cascade in visceral mesoderm precursors. Immunostaining of *jeb* mutant

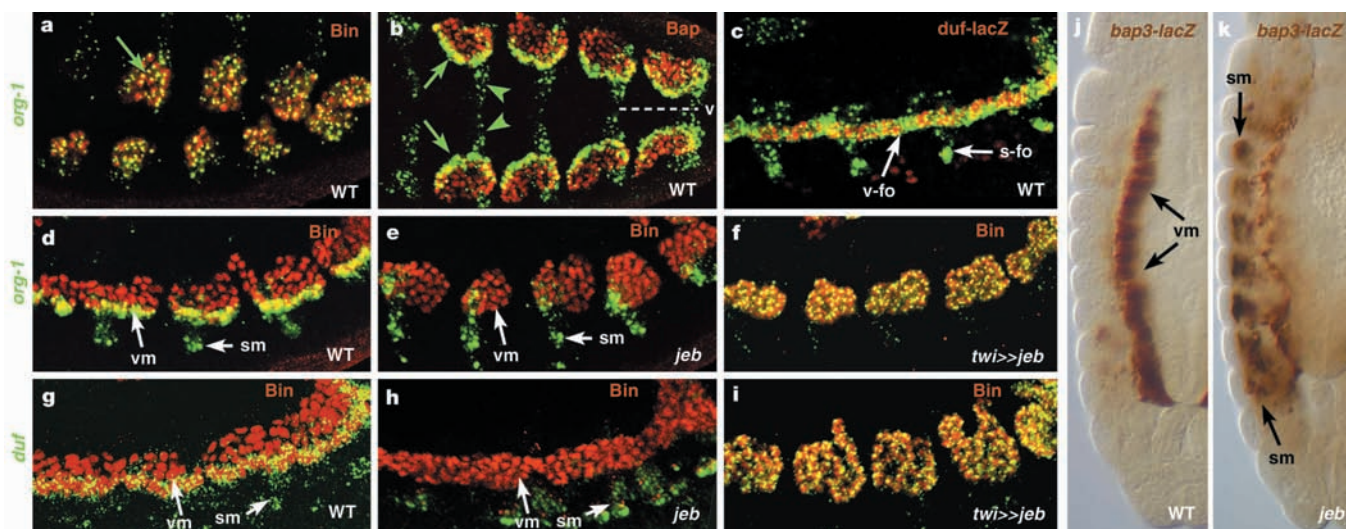


Figure 1 *jeb* specifies founder cell fates in the visceral mesoderm. Visceral mesoderm nuclei are marked by Bin or Bap protein staining (red), as indicated. mRNA probes are shown in green. **a**, Stage 9 embryo showing *org-1* mRNA (arrow) in all cells of the visceral mesoderm primordia. **b**, Ventral view of a stage 11 embryo. *org-1* expression (arrows) is restricted to the ventral row of cells in each visceral mesoderm cell cluster (arrowheads: somatic mesodermal *org-1*). v, ventral. **c–i**, Lateral high magnifications. **c**, At mid-stage 12, after a round of mitotic divisions of the progenitors, *org-1* mRNA (green) and *duf-lacZ* (*rP298* enhancer trap; red) are co-expressed in the two rows of visceral muscle founders. v-fo, visceral muscle founder; s-fo, somatic muscle founder. **d**, Stage 11 wild-type embryo showing *org-1* in muscle progenitors of the visceral mesoderm (vm). sm, somatic

mesoderm. **e**, Stage 11 *jeb* mutant embryo lacking *org-1* expression in the visceral mesoderm primordia. **f**, Stage 11 embryo; ectopic *jeb* expression in the early mesoderm results in ectopic *org-1* expression in all visceral mesoderm cells. **g**, Stage 12 wild-type embryo showing *duf* expression in visceral muscle founders. **h**, Stage 12 *jeb* mutant embryo showing lack of *duf* expression in the visceral mesoderm. **i**, Stage 12 embryo with *twi-driven* ectopic *jeb*, which results in *duf* expression in all cells of the visceral mesoderm primordia. **j**, Stage 13 *bap3-lacZ* wild-type embryo (ventral/left side) showing βGal in the visceral mesoderm. **k**, Analogous *jeb* mutant embryo carrying *bap3-lacZ* shows βGal predominantly in somatic muscle syncytia.

embryos demonstrates absent diphospho-MAPK in the ventral visceral mesoderm cells that normally accumulate Jeb and become founders (Fig. 2c). As with founder cell markers, ectopic Jeb produces ectopic diphospho-MAPK, but only in the visceral mesoderm (Fig. 2d).

The expanded expression of *org-1* upon mesodermal expression of activated versions of *Drosophila* Ras (Fig. 2e) and human Raf (Fig. 2f) implicates the Ras pathway in MAPK activation and founder cell specification in the visceral mesoderm. If Jeb signals through the Ras/MAPK pathway, then activation of this pathway should rescue *jeb* mutations. As shown in Fig. 2g, h, this prediction is true. As judged by expression of fasciclin III, a marker of visceral mesoderm differentiation, expression of activated Ras can substantially rescue *jeb* mutant embryos.

The observed effects of ectopic Jeb are limited to the visceral mesoderm. Together with the observation that uptake of Jeb into visceral mesoderm cells requires *shibire*-mediated endocytosis¹,

these data imply that Jeb acts through a tissue-specific receptor, which is coupled to the Ras/MAPK pathway. The receptor tyrosine kinase *Drosophila* *Alk*, a homologue of the human proto-oncogene anaplastic lymphoma kinase (ALK), is expressed in the early visceral mesoderm²¹. We therefore hypothesized that *Drosophila* *Alk* is the Jeb receptor. As shown in Fig. 3a, b, *Alk* messenger RNA is expressed in all cells of the trunk visceral mesoderm directly adjacent to the Jeb-expressing cells. In visceral mesoderm cells that both express *Alk* and take up Jeb¹ we detect diphospho-MAPK (Fig. 3a, b).

We tested whether, similar to Jeb, *Alk* activity is required for the specification of visceral mesoderm founder cells. As shown in Fig. 3d, embryos homozygous for a deficiency uncovering the *Alk* locus lack *org-1* expression in presumptive visceral mesoderm founders, a phenotype that can be rescued by expressing an *Alk* minigene in visceral mesoderm precursors (see Supplementary Fig. 2a, b). Mesodermal expression of a kinase-deficient, dominant interfering form of *Alk* produces an identical phenotype (Fig. 3e).

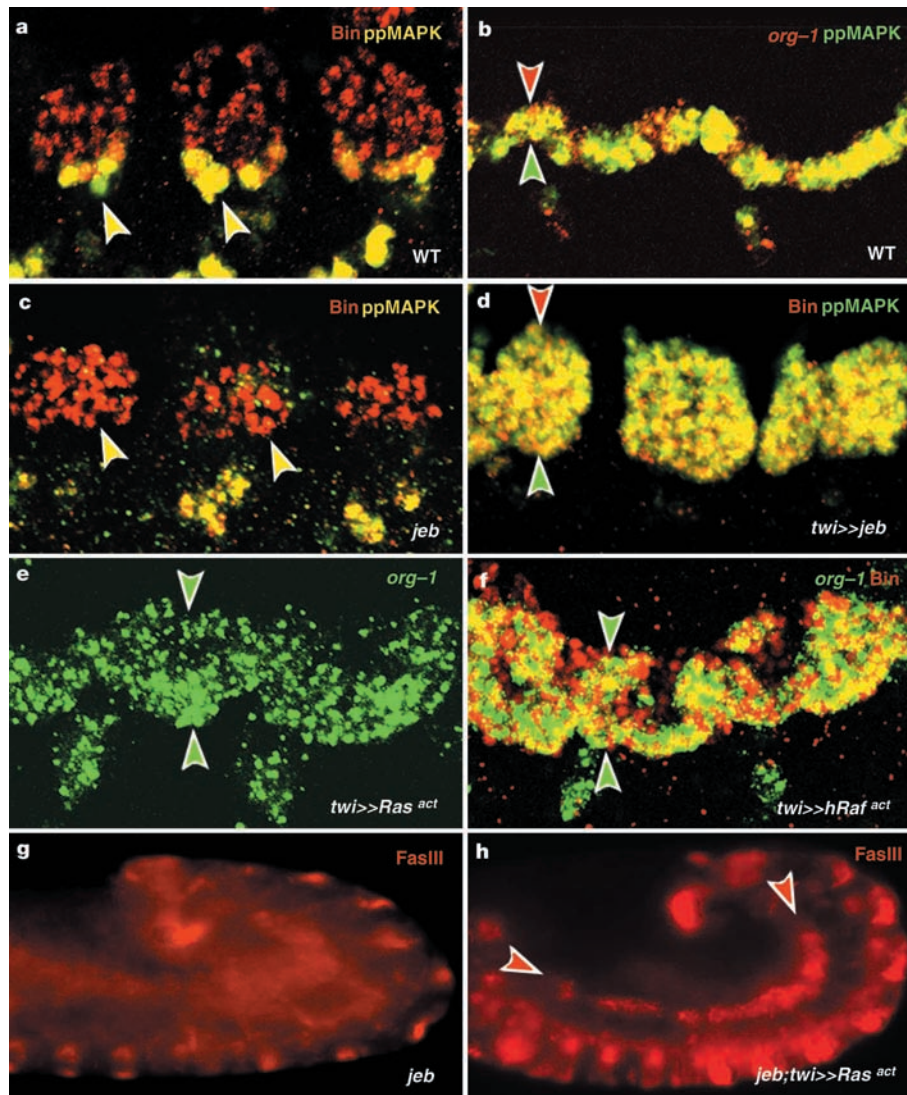


Figure 2 *jeb*-dependent Ras/MAPK signalling triggers founder cell specification in the visceral mesoderm. **a–f**, Lateral high magnifications. **a**, Anti-diphospho-MAPK (ppMAPK) antibodies detect activated MAPK (yellow, arrowheads) in the ventral cell rows of the visceral mesoderm primordia (red) at stage 10. **b**, Co-localization of ppMAPK (green, arrowhead) and *org-1* mRNA (red, arrowhead) at stage 11 identify cells with activated MAPK as visceral muscle founders. **c**, *jeb* mutant embryo lacking activated MAPK (yellow, arrowheads) in the visceral mesoderm primordia (red). **d**, Stage 11 embryo with *twi*-driven

ectopic *jeb* expression. All cells of the visceral mesoderm primordia (red, arrowhead) contain activated MAPK (green, arrowhead). **e**, **f**, Early stage 12 embryos with ectopic expression of activated Ras1 and activated human Raf showing strongly expanded *org-1* expression (green, arrowheads) in the visceral mesoderm. **g**, Stage 12 *jeb* mutant embryo showing loss of fasciclin III expression in the visceral mesoderm. **h**, Stage 12 *jeb* mutant embryo with ectopic expression of activated Ras1 as in **e**. Fasciclin III expression in the visceral mesoderm is rescued (arrowheads).

RNA-mediated interference (RNAi) injection experiments further confirm that *Alk* is specifically required for visceral mesoderm founder specification. β Gal staining of *bap3-lacZ* embryos injected with double-stranded (ds)*Alk* RNA demonstrates transformation of visceral into somatic muscle fates (Fig. 3f). Furthermore, injection of ds*Alk* RNA into *duf-lacZ* embryos results in strongly reduced or absent expression of this founder cell marker in the visceral mesoderm (Fig. 3h; see control in Fig. 3g). These RNAi phenotypes resemble the phenotypes of *jeb* mutant embryos, although they are less severe (see Fig. 1h, k).

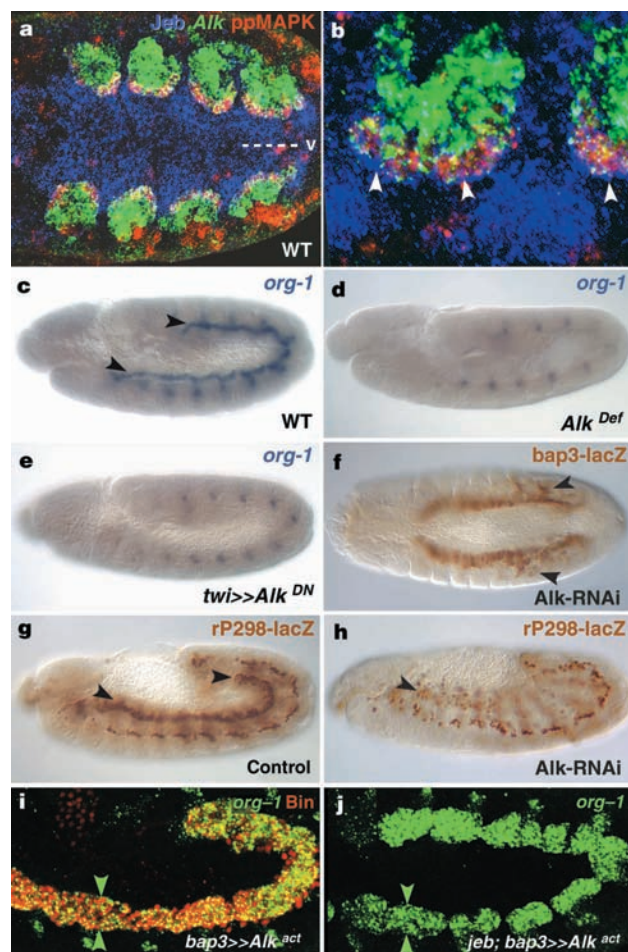


Figure 3 Jeb acts through the receptor tyrosine kinase Alk to specify visceral muscle founder cells. **a, b**, Ventral view (**a**) and high-magnification view of the lateral mesoderm (**b**) of stage 10 embryos triply stained for Jeb protein (blue), *Alk* mRNA (green) and activated MAPK (red). Jeb is predominantly detected in the ventrolateral somatic mesoderm and diffused Jeb is enriched along the visceral muscle progenitors (arrowheads), which contain activated MAPK. **c**, *org-1* mRNA expression in visceral (arrowheads) and somatic muscle progenitors of a stage 11 wild-type embryo. **d**, In a stage 11 embryo homozygous for *Df(2R)vg89e88* (*Alk*^{Def}) *org-1* expression in the visceral mesoderm is missing. **e**, Stage 11 embryo with ectopic expression of dominant-negative Alk lacking *org-1* expression in the visceral mesoderm. **f**, Late stage 12 *bap3-lacZ* embryo injected with ds*Alk* RNA. Near the posterior injection site, visceral mesoderm cells have assumed somatic mesodermal fates (arrowheads). **g**, Stage 12 *rP298* control embryo showing *duf* enhancer-driven β Gal in visceral (arrowheads) and somatic muscle founders. **h**, Stage 12 *rP298* embryo injected with ds*Alk* RNA posteriorly. Except for traces at the anterior (arrowhead), *duf* enhancer-driven β Gal is missing in the visceral mesoderm, but is unaffected in the somatic mesoderm. **i**, Stage 12 embryo with ectopic expression of activated Alk in the visceral mesoderm. *org-1* mRNA expression (green, arrowheads) has expanded into all cells of the trunk visceral mesoderm (red). **j**, Stage 11 *jeb* mutant embryo with ectopic expression of activated Alk in the visceral mesoderm. *org-1* mRNA expression in the visceral mesoderm (arrowheads) is rescued and expanded.

The loss of *duf* expression and expansion of *sns* expression in the visceral mesoderm on expression of dominant-negative Alk (see Supplementary Fig. 2c, d) is identical to a *jeb* null mutant phenotype as well. Conversely, the expansion of *org-1* expression in the visceral mesoderm on expression of activated Alk (a fusion protein analogous to the human oncogenic version, NPM-ALK²²) is indistinguishable from the effects of expression of ectopic Jeb, activated Ras and activated Raf (Fig. 3i; compare with Figs 1f, 2e and f, respectively). Finally, forced expression of activated Alk in homozygous *jeb* mutant backgrounds is able to rescue (and compared with wild type expand) *org-1* expression in the visceral mesoderm (Fig. 3j) and to restore midgut morphogenesis (see Supplementary Fig. 2e–g).

To confirm that Jeb signals through Alk we have determined that Jeb binds Alk with high affinity, and that Jeb binding to Alk activates the Ras/MAP kinase cascade. In these experiments we used Jeb–alkaline phosphatase fusion proteins (Jeb–AP)²³. To establish qualitatively the binding of Jeb to Alk, we visualized the specific association of Jeb–AP with *Alk*-transfected mammalian tissue culture cells. As shown in Fig. 4a, *Alk*-transfected cells bind Jeb–AP. By contrast, *Alk*-transfected cells do not bind either an equivalent concentration of alkaline phosphatase alone or a Jeb–AP fusion protein that lacks the type-A LDL receptor repeat in Jeb. This truncated version of Jeb resembles a mutant protein encoded by a null allele of *jeb*. The truncated protein does not accumulate in visceral mesoderm cells¹. Binding of Jeb depends on Alk, as demonstrated with non-transfected cells that were incubated with full-length Jeb–AP (data not shown).

We used a similar assay to demonstrate that the Jeb–Alk interaction is specific and has high affinity. As shown in Fig. 4b, Jeb binding to *Alk*-transfected cells is saturable at nanomolar concentrations. Scatchard analysis demonstrates a single class of high-affinity Jeb-binding site with a dissociation constant (K_d) of 2.2 nM (Fig. 4c). No binding was observed with either alkaline phosphatase alone or Jeb–AP that lacks the type-A LDL receptor repeat. We also confirmed Jeb-dependent activation of the Ras/MAP kinase cascade in this system (Fig. 4d). The concentration dependence of Ras/MAP kinase activation by Jeb correlates well with our binding data. Approximately half-maximal activation occurs in the range of 2–3 nM. As *in vivo*, removing the type-A LDL receptor repeat from Jeb abrogates Ras/MAP kinase activation (Fig. 4e).

We have shown that Jeb activates the Ras/MAPK cascade both *in vivo* and in *Alk*-transfected tissue culture cells. Jeb binds Alk with high affinity. *In vivo* Jeb accumulates in visceral muscle founder cells and, in late-stage embryos, in axons of the central nervous system. These patterns of Jeb accumulation are absent from *Alk*-deficient embryos (data not shown; see Supplementary Fig. 3) and in *jeb* mutants that produce an Alk-binding-deficient version of Jeb¹. Biochemical and genetic interference with Alk function produces phenotypes identical to *jeb* mutations. We have also shown that a critical function of Jeb signalling is to specify visceral muscle founder cells—patterning pioneers essential to midgut morphogenesis. Structurally Jeb belongs to a class of signalling molecules with type-A LDL receptor repeats as one of their functional domains. Others include *Caenorhabditis elegans* HEN-1 (ref. 24) and MIG-13 (ref. 25), and the mammalian proteins 8D6 (ref. 26) and scspondin²⁷. Jeb is the first among these to have an identified signalling receptor and a defined biological pathway. We anticipate that this discovery will lead to the identification of receptors and modes of action for other members of this class of signalling molecule.

The extracellular portions of mammalian and *Drosophila* Alk have common domain architectures. Their respective ligands are therefore also likely to share structural features. However, two closely related cytokines that are structurally unrelated to Jeb, pleiotrophin and midkine, have been identified by phage display as potential high-affinity ligands for human ALK²⁸. In *Drosophila*

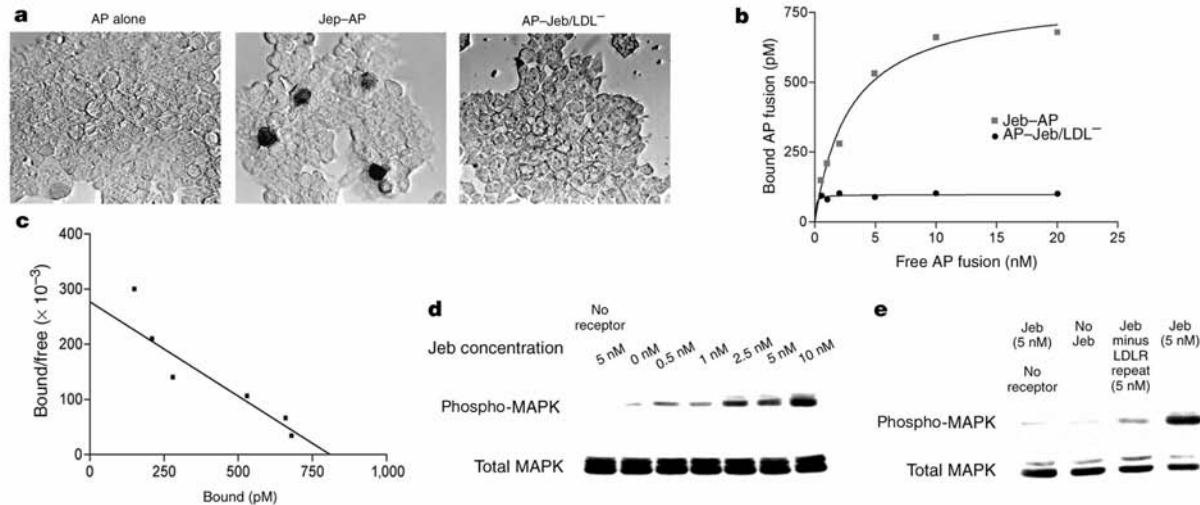


Figure 4 Jeb specifically binds with high affinity to *Drosophila* Alk and activates MAPK in mammalian tissue culture cells. **a**, Qualitative binding assays of Jeb-AP fusion protein to human 293T cells transfected with full-length *Alk* cDNA. Binding of AP is visualized by dark precipitates in transfected cells (middle panel). Cell-associated AP activity depends on Jeb and the type-A LDL receptor repeat in Jeb (right and left panels). **b**, Jeb binding to *Alk*-transfected 293T cells is saturable over a narrow concentration range and shows a K_d of 2.2 nM. *Alk*-transfected 293T cells were incubated with various concentrations of Jeb-AP and the retained AP activity was determined by quantitative colorimetric assays. As in

the qualitative binding assays, specific high-affinity binding is dependent on the type-A LDL receptor repeat in Jeb. **c**, Scatchard analysis demonstrates a single class of high-affinity binding sites on *Alk*-transfected 293T cells with a K_d of 2.2 nM. **d**, Increased concentration of Jeb-AP increases phospho-MAPK with half-maximal activation occurring at about 2 nM. **e**, Activation of the Ras/MAPK cascade, similar to binding, requires the type-A LDL receptor repeat in Jeb. Protein loading controls are shown as an immunoblot for total MAPK protein (**d**, **e**, bottom).

two clustered genes, *miple1* and *miple2*, encode polypeptides related to midkine/pleiotrophin. Similar to the mammalian genes, *Drosophila miple1* and *miple2* are expressed widely during embryogenesis (J.B.W., unpublished data). So, unlike Jeb, Miple1 and Miple2 cannot control the spatially restricted activation of Alk in the visceral mesoderm, although they may have an auxiliary function in Alk activation. The potential functions of Jeb-related molecules in mammalian Alk activation and the possible contribution of midkine/pleiotrophin-related factors to Alk signalling in *Drosophila* can now be tested by genetic and molecular approaches. The characterization of the Jeb/Alk signalling pathway in *Drosophila* is also likely to enhance our understanding of vertebrate Alk signalling in development and cancer. As most studies of mammalian Alk have focused on the role of oncogenic versions in cellular transformation, our current understanding of Alk's normal function in mammals is rudimentary (reviewed in ref. 2). In light of the known conservation of genetic pathways in the cardiac and splanchnic mesoderm^{8,29}, our insights into the regulation of *org-1* expression in *Drosophila* are potentially relevant for the understanding of the regulation of human *TBX1* and its roles in congenital cardiovascular and craniofacial disease. In addition, the specific expression of *Drosophila* and mouse Alk in the central nervous system^{2,21} suggests a conserved role of Alk signals in the development or function of neuronal tissues. □

Methods

Fly strains

We used the following *Drosophila* lines: *jeb*^{-c} (ref. 1), UAS-*jeb* (ref. 1), *Df(2R)Vg89e88* (an uninverted insertional transposition, *Tp(2;3)Vg89e88*, with a haplolethal deficiency component in the interval 52B3-C1; 53E2-F1), UAS-*Ras85D*^{V12}, UAS-*hRaf*, 2 × PE *twi-Gal4* (from the Bloomington Stock Collection), *rP298* (from A. Nose), *bap3-lacZ*, *bap3-Gal4* (ref. 8), UAS-*Alk*^{act}, UAS-*Alk*^{DN}, *bap3-Alk* (this study). Rescue of visceral mesodermal *Df(2R)Vg89e88* phenotypes was accomplished by recombining *bap3-Alk* onto the *Df(2R)Vg89e88* chromosome.

Plasmid constructions

The construct for constitutively active Alk (Alk^{act}) was made by fusing codons 1–117 of human nucleophosmin (NPM)²² to codons 1129–1701 of *Drosophila* Alk, cloned into

pUAST to produce UAS-*Alk*^{act} transgenic lines. The dominant negative Alk (Alk^{DN}) construct, also in pUAST, encodes the entire extracellular domain, transmembrane domain and a short tail of the intracellular domain (amino acids 1–1143) of Alk. The rescue construct, *bap3-Alk*, uses a *bap* enhancer element, *bap3* (ref. 8), and the *hsp70* basal promoter to drive the expression of *Alk* complementary DNA fused to an SV40 3' untranslated region. For details see Supplementary Information.

RNA interference and embryo staining

RNA interference was performed as described previously³⁰. *Alk* dsRNA (nucleotides 978–2089; 7.5 μM) was injected posteriorly into pre-blastoderm embryos from *bap3-lacZ* or *rP298* flies. Antibody and antibody/*in situ* hybridization double-fluorescent staining was performed as described previously^{3,8} (see Supplementary Information for antibodies and probes used).

AP fusion protein production

Full-length and truncated open reading frames of *jeb* were subcloned in frame with the APTag-4 expression vector (GenHunter). 293T cells, cultured with DMEM plus 10% serum in 100 mm plates, were transiently transfected for 24 h with the fusion constructs using Lipofectamine (Invitrogen). After 6–8 days of growth in Opti-MEM (Gibco) containing 0.3% serum or DMEM containing 0.3% serum, conditioned medium was concentrated by ultrafiltration. Stability of intact fusion protein was confirmed by immunoblotting using an anti-alkaline phosphatase antibody (GenHunter). The concentration of fusion protein was determined by measuring AP activity at 405 nm using a plate reader as described previously²³.

Cell staining and quantitative binding

293T cells were cultured on plates pre-coated with poly-D-lysine and transiently transfected for 24 h with Flag epitope-tagged Alk cDNA subcloned into the pCDNA3.1 expression vector using Lipofectamine. Binding of AP and AP fusions and staining was performed as described previously²³. For quantitative assays, 293T cells were transiently transfected for 24 h with pCDNA3.1-Alk. Cells were treated with a range of concentrations of Jeb-AP or AP-Jeb/LDL⁻ diluted in HBAH buffer (Hanks Balanced Salt Solution, BSA (0.5 mg ml⁻¹), 0.1% (w/v) NaN₃, 20 mM HEPES, pH 7.0) and binding to the cell surface was determined as described previously²³ using a plate reader to measure the change of absorbance (A) at 405 nm. After converting A to the quantity of protein bound, the binding curve and K_d were obtained using Prism3 software. Equivalent levels of full-length receptor expression in transfected cells were demonstrated by immunoblotting using anti-Flag M2 antibody (Sigma).

MAPK phosphorylation

After transfection of 293T cells as above, cells were serum-starved for 24 h in DMEM plus 0.3% serum, treated with 5 nM Jeb-AP or AP-Jeb/LDL⁻ produced in DMEM plus 0.3% serum for 20 min at 37 °C, placed on ice and washed twice with PBS. Cells were lysed with 0.5 ml lysis buffer (20 mM HEPES pH 7.5, 10 mM EGTA, pH 8, 1% NP-40, 2.5 mM MgCl₂, 200 mM sodium ortho-vanadate, 1 nM okadaic acid, protease inhibitor cocktail tablet

(Roche)) on ice. After clearing of the lysates by centrifugation, total and phosphorylated MAPK content was monitored by immunoblotting using p44/42 MAPK and phospho-p44/42 MAPK (Thr 202/Tyr 204) antibodies (New England Biolabs).

Received 23 April; accepted 14 July 2003; doi:10.1038/nature01916.

- Weiss, J. B., Suyama, K. L., Lee, H. H. & Scott, M. P. Jelly Belly: A *Drosophila* LDL receptor repeat-containing signal required for mesoderm migration and differentiation. *Cell* **107**, 387–398 (2001).
- Duyster, J., Bai, R. Y. & Morris, S. W. Translocations involving anaplastic lymphoma kinase (ALK). *Oncogene* **20**, 5623–5637 (2001).
- Azpiroz, N. & Frasch, M. *tinman* and *bagpipe*: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Dev.* **7**, 1325–1340 (1993).
- Bodmer, R. The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. *Development* **118**, 719–729 (1993).
- Frasch, M. Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early *Drosophila* embryo. *Nature* **374**, 464–467 (1995).
- Stachling-Hampton, K., Hoffmann, F. M., Baylies, M. K., Rushton, E. & Bate, M. *dpp* induces mesodermal gene expression in *Drosophila*. *Nature* **372**, 783–786 (1994).
- Azpiroz, N., Lawrence, P. A., Vincent, J.-P. & Frasch, M. Segmentation and specification of the *Drosophila* mesoderm. *Genes Dev.* **10**, 3183–3194 (1996).
- Zaffran, S., Küchler, A., Lee, H. H. & Frasch, M. *binou* (*FoxF*), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in *Drosophila*. *Genes Dev.* **15**, 2900–2915 (2001).
- Riechmann, V., Irion, U., Wilson, R., Grosskortenhaus, A. & Leptin, M. Control of cell fates and segmentation in the *Drosophila* mesoderm. *Development* **124**, 2915–2922 (1997).
- Lee, H. H. & Frasch, M. Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target *sloppy paired*. *Development* **127**, 5497–5508 (2000).
- Baylies, M. K. & Michelson, A. M. Invertebrate myogenesis: looking back to the future of muscle development. *Curr. Opin. Genet. Dev.* **11**, 431–439 (2001).
- Dworak, H. A. & Sink, H. Myoblast fusion in *Drosophila*. *Bioessays* **24**, 591–601 (2002).
- Bate, M. The embryonic development of larval muscles in *Drosophila*. *Development* **110**, 791–804 (1990).
- Dohrmann, C., Azpiroz, N. & Frasch, M. A new *Drosophila* homeobox gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis. *Genes Dev.* **4**, 2098–2111 (1990).
- San Martin, B., Ruiz-Gomez, M., Landgraf, M. & Bate, M. A distinct set of founders and fusion-competent myoblasts make visceral muscles in the *Drosophila* embryo. *Development* **128**, 3331–3338 (2001).
- Klapper, R. et al. The formation of syncytia within the visceral musculature of the *Drosophila* midgut is dependent on *dups* and *mbc*. *Mech. Dev.* **110**, 85–96 (2002).
- Porsch, M. et al. Isolation of a *Drosophila* T-box gene closely related to human *TBX1*. *Gene* **212**, 237–248 (1998).
- Baldini, A. DiGeorge syndrome: the use of model organisms to dissect complex genetics. *Hum. Mol. Genet.* **11**, 2363–2369 (2002).
- Gabay, L., Seger, R. & Shilo, B.-Z. MAP kinase in situ activation atlas during *Drosophila* embryogenesis. *Development* **124**, 3535–3541 (1997).
- Carmona, A., Gisselbrecht, S., Harrison, J., Jimenez, F. & Michelson, A. M. Combinatorial signaling codes for the progressive determination of cell fates in the *Drosophila* embryonic mesoderm. *Genes Dev.* **15**, 3910–3922 (1998).
- Loren, C. E. et al. Identification and characterization of DALK: a novel *Drosophila melanogaster* RTK which drives ERK activation *in vivo*. *Genes Cells* **6**, 531–544 (2001).
- Morris, S. W. et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263**, 1281–1284 (1994).
- Flanagan, J. G. & Cheng, H. J. Alkaline phosphatase fusion proteins for molecular characterization and cloning of receptors and their ligands. *Methods Enzymol.* **327**, 198–210 (2000).
- Ishihara, T. et al. HEN-1, a secretory protein with an LDL receptor motif, regulates sensory integration and learning in *Caenorhabditis elegans*. *Cell* **109**, 639–649 (2002).
- Sym, M., Robinson, N. & Kenyon, C. MIG-13 positions migrating cells along the anteroposterior body axis of *C. elegans*. *Cell* **98**, 25–36 (1999).
- Li, L. et al. Identification of a human follicular dendritic cell molecule that stimulates germinal center B cell growth. *J. Exp. Med.* **191**, 1077–1084 (2000).
- Gobron, S. et al. SCO-spondin is evolutionarily conserved in the central nervous system of the chordate phylum. *Neuroscience* **88**, 655–664 (1999).
- Stoica, G. E. et al. Identification of anaplastic lymphoma kinase as a receptor for the growth factor pleiotrophin. *J. Biol. Chem.* **276**, 16772–16779 (2001).
- Cripps, R. M. & Olson, E. N. Control of cardiac development by an evolutionarily conserved transcriptional network. *Dev. Biol.* **246**, 14–28 (2002).
- Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled2* act in the Wingless pathway. *Cell* **95**, 1017–1026 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Krauss, M. Mlodzik, H. Nguyen and R. Nusse for comments. We acknowledge the Developmental Studies Hybridoma Bank and the Bloomington Stock Collection for providing antibodies and fly stocks. The Mount Sinai Confocal Microscopy Shared Resource Facility was supported, in part, with funding from a NIH-NCI shared resources grant. This research was supported by grants from the National Institutes of Health to M.F. (from NICHD and NIDDK) and to J.B.W. (from NHLBI and HHMI).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.B.W. (weissjo@ohsu.edu) or M.F. (manfred.frasch@mssm.edu).

Jeb signals through the Alk receptor tyrosine kinase to drive visceral muscle fusion

Camilla Englund^{1*}, Christina E. Lorén^{1*}, Caroline Grabbe¹, Gaurav K. Varshney¹, Fabienne Deleuil^{1,2}, Bengt Hallberg² & Ruth H. Palmer¹

¹Umeå Center for Molecular Pathogenesis, ²Department of Medical Biosciences, Umeå University, Umeå, S-901 87, Sweden

*These authors contributed equally to this work

The *Drosophila melanogaster* gene *Anaplastic lymphoma kinase* (*Alk*) is homologous to mammalian Alk, a member of the Alk/Ltk family of receptor tyrosine kinases (RTKs)¹. We have previously shown that the *Drosophila* Alk RTK is crucial for visceral mesoderm development during early embryogenesis². Notably, observed *Alk* visceral mesoderm defects are highly reminiscent of the phenotype reported for the secreted molecule Jelly belly (*Jeb*)³. Here we show that *Drosophila* Alk is the receptor for Jeb in the developing visceral mesoderm, and that Jeb binding stimulates an Alk-driven, extracellular signal-regulated kinase-mediated signalling pathway, which results in the expression of the downstream gene *duf* (also known as *kirre*)^{4,5}—needed for muscle fusion. This new signal transduction pathway drives specification of the muscle founder cells, and the regulation of *Duf* expression by the *Drosophila* Alk RTK explains the visceral-mesoderm-specific muscle fusion defects observed in both *Alk* and *jeb* mutant animals.

Alk was first described in non-Hodgkin's lymphoma^{6,7}, and is now known to be involved in many genetic translocation events^{8,9}. Alk function in higher vertebrates has remained elusive, despite identification of mouse and human homologues^{10,11}. *Drosophila* Alk is expressed in the developing visceral mesoderm during embryogenesis¹, and *Alk* mutants reveal that no functional midgut is formed². The visceral musculature in *Drosophila* is syncytial^{12,13}, and arises during embryogenesis through the fusion of multiple myoblasts^{14,15}. The process of myoblast fusion in *Drosophila* is a dynamic relationship between two myoblast types: the 'founder cells' and the 'fusion-competent myoblasts'¹⁶. Progenitor myoblasts give rise to founder cells that serve as 'seeds' for muscle formation, which the fusion-competent myoblasts then recognize and fuse with^{16,17}.

Mutant *Drosophila* *Alk* and *jeb* larvae do not ingest food (Fig. 1; compare b, c with a) and lack discernible intestinal structures, whereas heterozygous siblings are robust with healthy appetites. Therefore, embryonic visceral mesoderm development was further analysed using fasciclin III, which is a marker for differentiated visceral mesoderm. In wild-type embryos Alk and fasciclin III expression patterns overlap as the midgut takes form (Fig. 1d, g; j, m in detail). Alk mutant protein is visible in *Alk* mutants because our Alk antibodies are raised to an amino-terminal epitope². In *Alk* mutant embryos the visceral mesoderm is observed as a disorganized group of Alk-positive cells at stage 13 (Fig. 1; compare e with d; panel o in detail)². A similar phenotype is observed in *jeb* mutants, with Alk-positive visceral mesoderm cells scattered in a disorganized manner instead of the organized band of cells normally observed at this stage (Fig. 1; compare f with d; panel n in detail). The earliest developmental stage at which a mutant phenotype in the visceral mesoderm can clearly be observed in *jeb* and *Alk* mutants is at stage 11 (Fig. 1j–l). In wild type the muscle founder cells are specified in the visceral mesoderm, and become arranged in an organized column of cells located ventrally (Fig. 1j,

arrowheads indicate the column of muscle founder cells). In *Alk* and *jeb* mutant embryos this organized column of cells is never observed (Fig. 1k, l), although the cells within the clusters are still capable of migrating in an anterior and posterior direction to create a semi-continuous band of visceral mesoderm. Thus, the first obvious visceral mesoderm defect to be observed in both *jeb* and *Alk* mutants is a lack of muscle founder cells.

During stage 11 of embryogenesis activation of extracellular signal-regulated kinase (ERK) can be observed in the visceral

mesoderm^{1,18}. ERK activation is observed specifically in the muscle founder cells and not the fusion-competent cells (Fig. 2a, b). *Drosophila* *Alk* is the RTK responsible for this activation of ERK, as in *Alk* mutants there is a complete absence of diphospho-ERK in the visceral arches at stage 11 (ref. 2) (compare Fig. 2c with a). Furthermore, diphospho-ERK is also absent from the visceral arches at stage 11 in *jeb* mutants (Fig. 2; compare e with a), implying that *Jeb* function is also required for ERK activation in the visceral mesoderm. Thus, we conclude that both the secreted molecule *Jeb* and the *Alk* RTK are critical for ERK activation in the muscle founder cells of the visceral mesoderm at stage 11.

As the *Duf* protein has an important role in myoblast aggregation and fusion^{4,5}, we decided to explore the possibility that an *Alk*-stimulated, ERK-mediated signalling pathway could drive *Duf* expression. Notably, *Duf* is expressed by the founder cells of the visceral mesoderm and has been shown to have an attractive function^{4,5}. In *Alk* mutant embryos we observed that *Alk* is indeed required for expression of *duf* in the muscle founder cells of the visceral mesoderm at stage 11 (compare Fig. 2h with g). As one would predict for a putative *Drosophila* *Alk* ligand, *duf* expression was also abrogated in the visceral mesoderm muscle founder cells of *jeb* mutant embryos (compare Fig. 2j with i). Expression of *duf* in

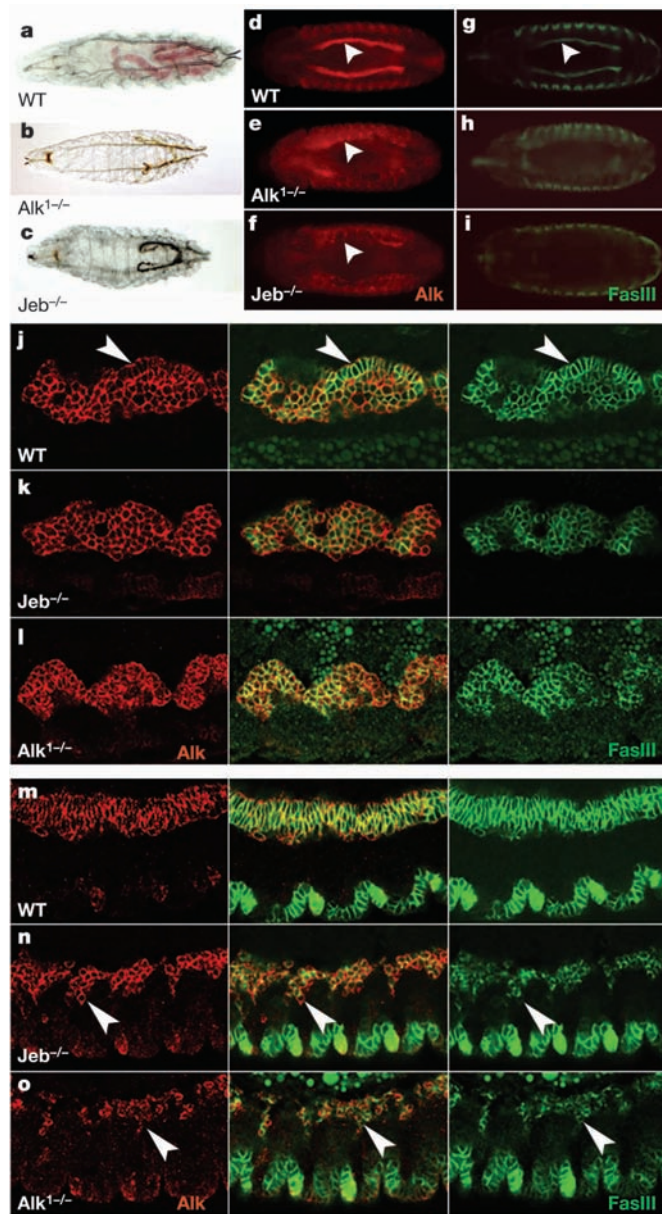


Figure 1 *Jeb* and *Alk* mutants show identical visceral mesoderm phenotypes. **a–c**, Larval eating assay. Wild type (**a**), *Alk* (**b**) and *jeb* (**c**) first instar mutant animals are shown. **d–i**, Visceral mesoderm at stage 14, dorsal view. *Alk* (red) and fasciclin III (green) staining is shown. **d, g**, Wild-type; **e, h**, *Alk*; **f, i**, *jeb* mutant animals. Embryos are oriented with anterior to left. **j–l**, Confocal analysis of stage 11 visceral mesoderm. Wild-type embryos display an organized row of columnar muscle founder cells (**j**, arrowheads), which is absent in *jeb* (**k**) and *Alk* (**l**) mutant animals. **m–o**, Visceral mesoderm at stage 13. **m**, Wild type; **n**, *jeb*; **o**, *Alk* mutant animals. Arrowheads in **n** and **o** indicate disorganized clusters of visceral mesoderm cells in *jeb* and *Alk* mutant embryos. *Alk* (red) and anti-fasciclin III (green) staining is shown.

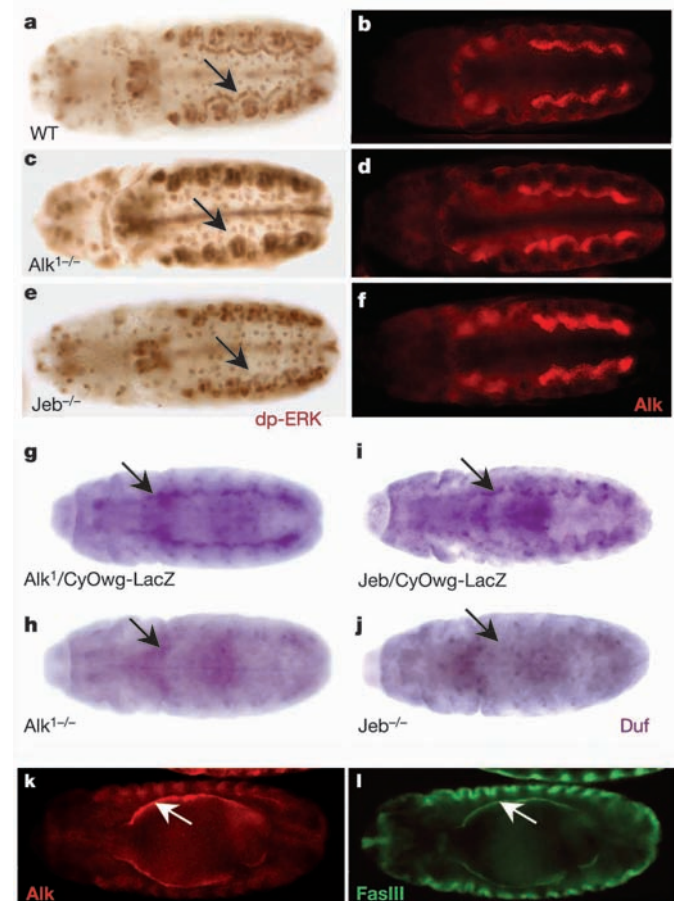


Figure 2 *Jeb* and *Alk* are both required for activation of ERK and expression of *Duf* in muscle founder cells. **a–f**, ERK activation in stage 11 visceral mesoderm. Diphospho (dp)-ERK (brown) and *Alk* (red) staining is shown. **a, b**, Wild type; **c, d**, *Alk*^{1-/-}; **e, f**, *jeb* mutant animals. ERK activation is detected in visceral mesoderm muscle founder cells in wild-type embryos (arrows) but not in *Alk*^{1-/-} or *jeb* mutant embryos. **g–j**, *duf* expression in muscle founder cells of the visceral mesoderm at stage 11 requires both *Jeb* and *Alk* function. *duf* *in situ* analysis of heterozygous sibling (**g, i**), *Alk*^{1-/-} mutant (**h**) and *jeb* mutant (**j**) embryos are shown. **k, l**, *jeb* mutants are rescued by ectopic expression of *Alk*. *Alk*E16.5-Gal4 driving UAS-*Alk* results in a rescue of fasciclin-III-positive gut structures in *jeb* mutant embryos (**l**, arrow).

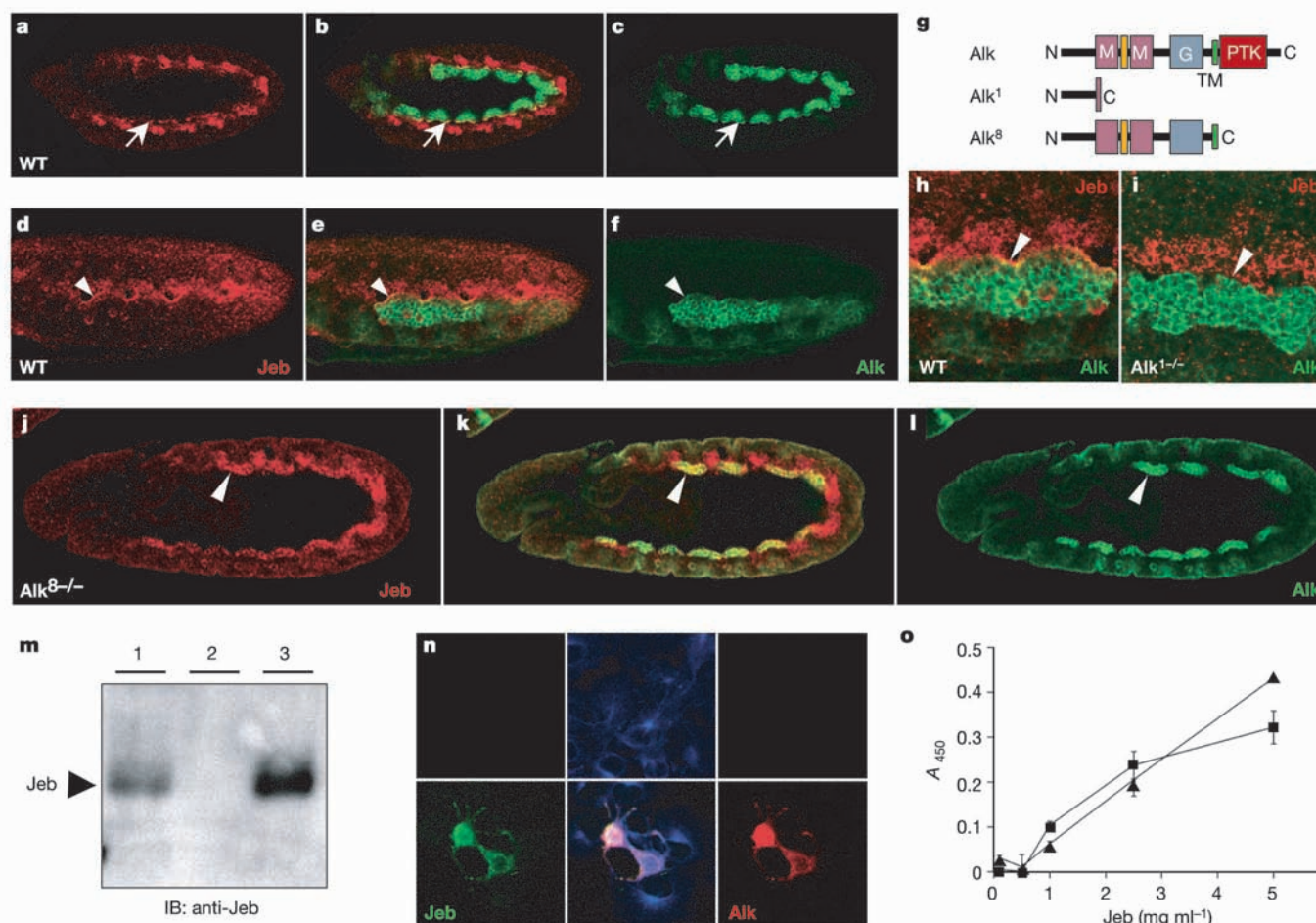


Figure 3 Jeb is taken up by Alk in the visceral mesoderm. **a–f**, Confocal analysis of Jeb uptake by the visceral mesoderm in wild-type embryos. Jeb (red) and Alk (green) staining is shown. **a–c**, Stage 10 embryos, lateral view. **d–f**, Stage 11 embryos, dorso-lateral view. Arrows indicate co-localization of Jeb and Alk. **g**, Schematic representation of the Alk mutant alleles *Alk*¹ and *Alk*⁸. The MAM (M), glycine-rich (G), transmembrane (TM) and PTK domains of Alk are indicated. **h, i**, Jeb is bound and taken up by the muscle founder cells in wild-type embryos (**h**, arrowhead), but not in *Alk*¹ mutant embryos (**i**, arrowhead); stage 11, dorso-lateral view. **j–l**, Jeb is bound nonspecifically to all cells of the developing visceral mesoderm in *Alk*⁸ mutant embryos (arrowheads); stage 10, lateral view. **m–o**, Jeb binds directly to Alk. **m**, Jeb co-immunoprecipitates with Alk. Endogenous Alk was immunoprecipitated with anti-Alk antibodies and immunoblotted for

Jeb (lane 1). Ectopically expressed Alk cytoplasmic domain (Twist-Gal4; UAS-*Alk*^A (lane 2)) and full-length Alk (Twist-Gal4; UAS-*Alk*^{fl} (lane 3)) were immunoprecipitated and immunoblotted for Jeb. **n**, Recombinant Jeb binds to Alk-expressing cells. COS7 cells were transfected with Alk (bottom row) or control (top row) before addition of purified Jeb protein. Alk-expressing cells specifically bind Jeb protein (bottom row). Jeb (green), Alk (red) and tubulin (blue) were used to visualize cells. **o**, ELISA analysis of Jeb–Alk binding. Alk exhibited a high affinity for purified recombinant Jeb by direct ELISA analysis (triangles). Jeb exhibited a high affinity for recombinant Alk by sandwich ELISA analysis (squares). Values represent an average of three experiments. A_{450} , absorbance at 450 nm.

the somatic musculature is unaffected, a result one would anticipate as Alk is not expressed in this tissue (Supplementary Fig. 1). Additionally, *duf* expression in the central nervous system is unaffected (Supplementary Fig. 1). Thus, we conclude that Alk and Jeb are required not only for activation of a diphospho-ERK-mediated signalling pathway in the visceral mesoderm, but also for the expression of the Duf protein, which is known to be important for the muscle fusion process.

If Alk is the receptor for Jeb then overexpression of Alk should be capable of rescuing the visceral mesoderm phenotypes observed in a *jeb* mutant. By stage 15 no fasciclin-III-positive gut structures are observed in *jeb* mutant animals. Ectopic expression of Alk in the visceral mesoderm results in a partial rescue of fasciclin-III-positive gut structures (Fig. 2l, arrow, compare with Fig. 1i). Thus, it seems that Alk is genetically downstream of Jeb function, supporting the hypothesis that Alk is indeed the receptor for the secreted Jeb protein in the visceral mesoderm.

The Jeb protein is a new extracellular signalling molecule that is

transcribed in the neighbouring somatic mesoderm, and is then specifically taken up by the visceral mesoderm cells³, suggesting that a tissue-specific receptor mediates the uptake of Jeb. If Alk functions as a receptor for Jeb, then Jeb uptake by the visceral mesoderm should be affected in *Alk* mutants. Analysis of Jeb and Alk localization in the visceral mesoderm reveals a complex interplay between Jeb and the Alk-positive visceral mesoderm (Fig. 3). In wild-type embryos Jeb is expressed in cells ventral to the Alk-positive visceral mesoderm clusters, and at the sites of contact between these cells a clear co-localization of Jeb and Alk can be observed (Fig. 3a–c, arrows). On closer inspection, the Alk-positive visceral mesoderm cells, which contact the Jeb-secreting mesoderm, appear to respond to Jeb by differentiating into an organized column of muscle founder cells. By stage 11, Jeb protein can be observed in a wavy line reflecting the organized column of muscle founder cells (Fig. 3d–f, arrowheads). In both *Alk* and *jeb* mutant animals the organized column of muscle founder cells is never observed (Fig. 1k, l), thus raising the possibility that the Alk-

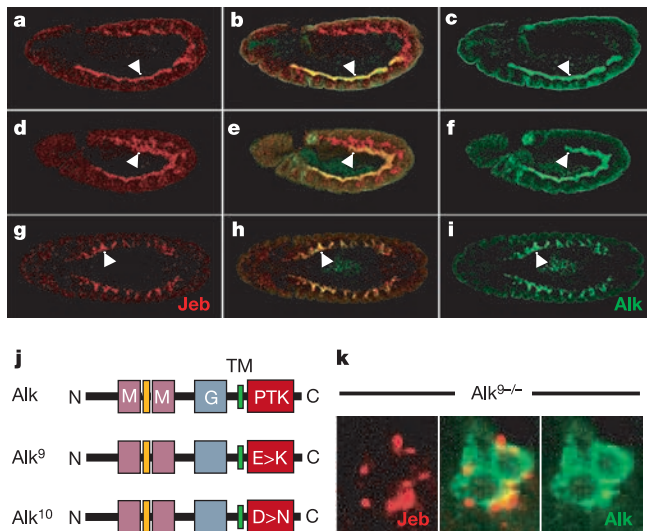


Figure 4 Alk kinase activity is required for the uptake of Jeb in the visceral mesoderm. **a–i**, Confocal analysis showing that Jeb is irreversibly bound but not internalized by all Alk-positive visceral mesoderm cells in *Alk⁹* and *Alk¹⁰* mutant embryos (*Alk⁹* is shown). Binding of Jeb to the visceral mesoderm persists through embryogenesis, so that Jeb is still visualized even at later stages (**g–i**, arrowheads). **a–c**, Stage 11 embryos, lateral view; **d–f**, stage 12 embryos, lateral view; **g–i**, stage 13 embryos, dorsal view. **j**, Schematic representation of the Alk mutant alleles *Alk⁹* and *Alk¹⁰* (both mutated in the kinase domain)². **k**, Confocal analysis showing that Jeb protein is not internalized and degraded by visceral mesoderm cells in *Alk⁹* and *Alk¹⁰* mutant embryos. Jeb (left, red), Alk (right, green) and merged images (middle, yellow) are shown.

positive visceral mesoderm responds spatially to the Jeb signal from the adjacent tissue to specify an organized column of muscle founder cells in the correct orientation. In *Alk¹* mutant flies, which express a short N-terminal fragment of the Alk RTK containing no detectable functional domains² (Fig. 3g), we observe no co-localization or uptake of Jeb in the visceral mesoderm, in agreement with Alk functioning as a receptor for Jeb (compare Fig. 3i with h; arrowheads). Furthermore, in *Alk⁸* mutants, in which the Alk protein is truncated just after the transmembrane domain but still has an intact Alk extracellular domain (Fig. 3g), we observe that Jeb protein is tightly bound by all *Alk⁸*-positive cells of the visceral mesoderm (Fig. 3j–l; arrowheads). This suggests that Alk is required not only for the binding and signal transduction response to Jeb, but also for degradation of the Jeb–Alk complex. Furthermore, we find that Jeb co-immunoprecipitates with the endogenous Alk RTK (Fig. 3m, lane 1), and this is dependent on the extracellular domain of Alk (Fig. 3m, lanes 2, 3). Recombinant Jeb protein binds specifically to cells overexpressing Alk (Fig. 3n), and this is dependent on the extracellular portion of Alk, as overexpression of the extracellular domain alone is sufficient for Jeb binding (data not shown). Finally, a direct binding of Jeb and Alk was observed, both by direct enzyme-linked immunosorbent assay (ELISA) analysis and by sandwich ELISA (Fig. 3o, triangles and squares, respectively). Thus, Alk seems to function as the receptor, or part of a receptor complex, for Jeb.

To investigate the role of Alk protein tyrosine kinase (PTK) activity *in vivo* we used the Alk PTK mutants *Alk⁹* and *Alk¹⁰*, which each contain a point mutation in vital amino acid residues within the PTK domain (Fig. 4j)². Analysis of Jeb localization in *Alk⁹* and *Alk¹⁰* mutants shows that Jeb protein is bound by the mutant receptors as would be predicted, but is not taken up by the visceral mesoderm cells (Fig. 4a–i). Thus, the PTK activity of Alk, although not required for binding of Jeb, is required for the uptake into cells

and subsequent degradation of Jeb in the visceral mesoderm (Fig. 4k). Furthermore, this Jeb–Alk interaction seems to be stable as Jeb protein can still be visualized at late stage 13, long after Jeb protein ceases to be present in wild-type embryos (Fig. 4g–i). Thus, the PTK activity of the Alk RTK seems to be crucial for the uptake of Jeb by the visceral mesoderm.

We have examined the fate of the visceral mesoderm cells in *jeb* and *Alk* mutants at later stages (Supplementary Fig. 2), and find that these mesoderm cells are still fusion competent. Indeed, it seems to be the absence of the muscle founder cells with which to seed the fusion process in the visceral musculature that leads to the loss of visceral muscle in these mutants. Thus, *jeb* or *Alk* mutant fusion-competent myoblasts derived from the visceral mesoderm are still able to incorporate into muscle, albeit somatic muscle, in what seems to be a competitive manner.

We have shown that the *Drosophila* Alk RTK is the receptor, or part of a receptor complex, responsible for binding the recently identified Jeb protein, which is required for visceral mesoderm migration and differentiation³. Jeb and Alk signal through an ERK-mediated signalling pathway to specify the visceral mesoderm muscle founder cells. In the absence of either Jeb or Alk function there is a critical failure in the fusion process within the visceral mesoderm. The targets of Jeb-induced Alk-mediated signalling include the fusion determinant Duf^{4,5}, thus offering an explanation for the muscle-fusion defects observed in *jeb* and *Alk* mutant flies. Further exploration of Jeb/Alk-mediated signalling pathways and targets is an important task for the future. □

Methods

Fly culture

We used standard *Drosophila* husbandry procedures. Flies were raised and crossed at room temperature unless otherwise stated. The Jeb mutant fly line (Bloomington 10576) has previously been described³. *Drosophila* Alk mutant alleles, *Alk^{6.5}-Gal4*, and the UAS-*Alk^{fl}* transgenic fly line have previously been described^{1,2}. UAS-*Alk^{act}* is a ligand-independent constitutively active Alk transgenic fly line. CyOKr-GFP and CyWg-LacZ balancers were used to genotype embryos.

Larval eating assay

jeb and *Alk* mutant fly lines were balanced over CyOKr-GFP to allow identification of mutant first instar larvae. Embryos were collected and raised on food that had been precoloured with food dye. *jeb* and *Alk* mutant larvae were identified by the absence of Kr-GFP expression.

Immunostaining

Embryos were fixed and immunostained as described¹⁹. The following primary antibodies were used: rabbit anti-β-galactosidase (diluted 1:1500; Cappel), mouse anti-fasciclin III (monoclonal antibody 7G10, 1:50; Developmental Studies Hybridoma Bank), rabbit anti-Mef2 (1:500; provided by B. Paterson), mouse anti-MHC (1:5), mouse anti-phospho-ERK (1:500, Sigma), mouse anti-α-tubulin (1:250 dilution), guinea-pig anti-Alk and rabbit anti-Alk (both 1:1,000; ref. 2). The following secondary antibodies were used: biotin-conjugated goat anti-mouse IgG (1:1,000), Cy3-conjugated donkey anti-guinea pig (1:200), Cy5-conjugated donkey anti-mouse (1:200), Cy2-conjugated goat anti-mouse IgG (1:1,000, all four from Jackson Laboratories), Cy2-conjugated goat anti-rabbit IgG (1:1,000), Cy3-conjugated goat anti-rabbit IgG (1:1,000, both from Amersham). Signal was developed using Vectastain Elite ABC Kit (Vector Laboratories) and DAB reaction. Embryos were cleared in methyl salicylate (Sigma) before visualization. Embryo staging was carried out according to ref. 20.

Anti-Jeb antibodies

His–Jeb fusion protein encoding amino acids 122–1684 of the Jeb protein was generated in pETM-11. The orientation and reading frame of the His–Jeb sequence was subsequently confirmed by DNA sequence analysis. His–Jeb fusion protein was induced and purified from *Escherichia coli* (BL21(DE3)) bacterial lysates by standard protocols using Ni-NTA agarose (Qiagen). The resulting eluted His–Jeb recombinant protein was used for guinea-pig immunization. Polyclonal anti-Jeb antibodies were used at 1:1,000 with signal amplification by incubation with Vectastain Elite ABC Kit (Vector Laboratories) and Cy3 tyramide signal amplification (NEN).

Immunoprecipitation and immunoblotting analysis

Standard protocols for protein extraction, immunoprecipitation and immunoblot analysis were used. In brief, embryos were lysed in RIPA buffer²⁰ and the resulting cell lysates were immunoprecipitated with antibodies, before SDS–PAGE and immunoblotting with anti-Jeb antibodies. Endogenous *Drosophila* Alk was immunoprecipitated with anti-Alk antibodies. Ectopically expressed Alk cytoplasmic domain (Twist-Gal4; UAS-*Alk^{act}*) was immunoprecipitated with anti-

haemagglutinin antibodies, and ectopically expressed full-length (fl) Alk (Twist-Gal4; UAS-Alk^{fl}) was immunoprecipitated with anti-Alk antibodies.

ELISA and immunofluorescence analysis

ELISA was carried out according to standard protocols. Microtitre plate wells were coated with *Drosophila* Alk monoclonal antibody 123 (ref. 1) for the sandwich ELISA, and with purified recombinant His-Jeb protein for the direct ELISA. For immunofluorescence staining, COS7 cells were transfected with either pcDNA3 (as control), pcDNA3-Alk^{fl}, pcDNA3-Alk^E, or pcDNA3-Alk^A (where superscript E and A indicated extracellular and activated, respectively), and purified recombinant His-Jeb protein was added (to a final concentration of 1 µg ml⁻¹) for 1 h before staining, as described previously²¹. Analysis was carried out by confocal laser scanning microscopy (Leica).

In situ hybridization

In situ hybridization of whole-mount embryos was done with digoxigenin-labelled *duf* RNA as probe and followed by antibody staining as described²².

Received 8 June; accepted 29 July 2003; doi:10.1038/nature01950.

- Loren, C. E. *et al.* Identification and characterization of DALK: a novel *Drosophila melanogaster* RTK which drives ERK activation *in vivo*. *Genes Cells* **6**, 531–544 (2001).
- Loren, C. E. *et al.* A crucial role for the Anaplastic lymphoma kinase receptor tyrosine kinase in gut development in *Drosophila melanogaster*. *EMBO Rep.* **4**, 781–786 (2003).
- Weiss, J. B., Suyama, K. L., Lee, H. H. & Scott, M. P. Jelly belly: a *Drosophila* LDL receptor repeat-containing signal required for mesoderm migration and differentiation. *Cell* **107**, 387–398 (2001).
- Ruiz-Gomez, M., Coutts, N., Price, A., Taylor, M. V. & Bate, M. *Drosophila* dumbfounded: a myoblast attractant essential for fusion. *Cell* **102**, 189–198 (2000).
- Strunkelberg, M. *et al.* rdt and its paralogue kirre act redundantly during embryonic muscle development in *Drosophila*. *Development* **128**, 4229–4239 (2001).
- Fujimoto, J. *et al.* Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proc. Natl Acad. Sci. USA* **93**, 4181–4186 (1996).
- Morris, S. W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263**, 1281–1284 (1994).
- Falini, B. Anaplastic large cell lymphoma: pathological, molecular and clinical features. *Br. J. Haematol.* **114**, 741–760 (2001).
- Duyster, J., Bai, R. Y. & Morris, S. W. Translocations involving anaplastic lymphoma kinase (ALK). *Oncogene* **20**, 5623–5637 (2001).
- Morris, S. W. *et al.* ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene* **14**, 2175–2188 (1997).
- Iwahara, T. *et al.* Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* **14**, 439–449 (1997).
- Klapper, R. *et al.* The formation of syncytia within the visceral musculature of the *Drosophila* midgut is dependent on *duf*, *sns* and *mbc*. *Mech. Dev.* **110**, 85–96 (2002).
- Martin, B. S., Ruiz-Gomez, M., Landgraf, M. & Bate, M. A distinct set of founders and fusion-competent myoblasts make visceral muscles in the *Drosophila* embryo. *Development* **128**, 3331–3338 (2001).
- Dworak, H. A. & Sink, H. Myoblast fusion in *Drosophila*. *Bioessays* **24**, 591–601 (2002).
- Baylies, M. K. & Michelson, A. M. Invertebrate myogenesis: looking back to the future of muscle development. *Curr. Opin. Genet. Dev.* **11**, 431–439 (2001).
- Bate, M. The embryonic development of larval muscles in *Drosophila*. *Development* **110**, 791–804 (1990).
- Carmena, A., Bate, M. & Jimenez, F. Lethal of scute, a proneural gene, participates in the specification of muscle progenitors during *Drosophila* embryogenesis. *Genes Dev.* **9**, 2373–2383 (1995).
- Gabay, L., Seger, R. & Shilo, B. Z. MAP kinase in situ activation atlas during *Drosophila* embryogenesis. *Development* **124**, 3535–3541 (1997).
- Patel, N. H. Imaging neuronal subsets and other cell types in whole-mount *Drosophila* embryos and larvae using antibody probes. *Methods Cell Biol.* **44**, 445–487 (1994).
- Campos-Ortega, J. A. & Hartenstein, V. The Embryonic Development of *Drosophila Melanogaster*. (Springer, Berlin, 1997).
- Palmer, R. H. *et al.* DFak56 is a novel *Drosophila melanogaster* focal adhesion kinase. *J. Biol. Chem.* **274**, 35621–35629 (1999).
- Jiang, W. *et al.* PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis. *Mol. Cell* **2**, 877–885 (1998).
- Kopczynski, C. C., Davis, G. W. & Goodman, C. S. A neural tetraspanin, encoded by late bloomer, that facilitates synapse formation. *Science* **271**, 1867–1870 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors would like to thank T. Hunter and I. Salecker for critical reading of the manuscript, and D. Eriksson, N. Norgren and A. Sheikholvaezin for help with ELISA analysis. This work is funded by the Swedish Research Council and is also supported by The Swedish Society for Medical Research (SSMF), Åke Wibergs Fund, the Royal Swedish Academy of Sciences, Lars Hiertas Minne Fund, and the Cancer Research Fund of Northern Sweden. F.D. was supported by a post-doctoral fellowship from the Wenner-Grenska Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.H.P. (Ruth.Palmer@ucmp.umu.se).

Molecular identification of a danger signal that alerts the immune system to dying cells

Yan Shi¹, James E. Evans² & Kenneth L. Rock¹

¹Department of Pathology, and ²Proteomics and Mass Spectrometry Facility, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA

In infections, microbial components provide signals that alert the immune system to danger and promote the generation of immunity^{1,2}. In the absence of such signals, there is often no immune response or tolerance may develop. This has led to the concept that the immune system responds only to antigens perceived to be associated with a dangerous situation such as infection^{3,4}. Danger signals are thought to act by stimulating dendritic cells to mature so that they can present foreign antigens and stimulate T lymphocytes^{2,5–7}. Dying mammalian cells have also been found to release danger signals of unknown identity^{8–11}. Here we show that uric acid is a principal endogenous danger signal released from injured cells. Uric acid stimulates dendritic cell maturation and, when co-injected with antigen *in vivo*, significantly enhances the generation of responses from CD8⁺ T cells. Eliminating uric acid *in vivo* inhibits the immune response to antigens associated with injured cells, but not to antigens presented by activated dendritic cells. Our findings provide a molecular link between cell injury and immunity and have important implications for vaccines, autoimmunity and inflammation.

When dying cells are co-injected with antigen into animals, they provide an adjuvant effect for priming T-cell responses^{10,11}. This endogenous adjuvant activity is present in the cytosol of cells and markedly increases when cells are injured, for example, by ultraviolet irradiation. To identify this endogenous adjuvant, here we fractionated cytosol from ultraviolet-irradiated BALB/c 3T3 cells by high-performance liquid chromatography (HPLC) on a sizing column monitored with a diode array ultraviolet spectrum detector (Fig. 1a). Pools of 4–5 consecutive fractions were tested for their ability to augment the priming of CD8⁺ T-cell responses when co-injected with particulate HIV gp120 antigen. After 14 d, splenocytes from the primed mice were stimulated *ex vivo* with antigen and then assayed for their ability to kill antigen-bearing target cells in a ⁵¹Cr-release assay¹⁰.

A pool of low molecular weight (LMW) fractions that were below the optimal resolution range (relative molecular mass less than 5,000; *M_r* < 5K) of the sizing columns had adjuvant activity that markedly enhanced the generation of cytotoxic T lymphocyte (CTL) responses (Fig. 1b). When the individual components of this active LMW pool were tested, most of the activity was located in a single fraction (Fig. 1b, inset). Another pool of higher molecular weight fractions (~40–100K) also had adjuvant activity, but this has not been further studied at this time.

We focused on characterizing the LMW adjuvant. It was present in the cytosol of both ultraviolet-treated 3T3 cells and liver. LMW fractions from liver cytosol were separated further by HPLC with sequential anion exchange (Mono Q HR5/5; Fig. 1c), sizing (Superdex 200 HR 10/30; Fig. 1d) and reverse phase (C18) (Fig. 1e) columns. After each separation the active fractions were identified by their ability to boost CTL responses *in vivo* (Fig. 1f–h). On all four columns, the LMW adjuvant from liver showed the same chromatography profile as that from 3T3 cells, suggesting that both cells contained the same adjuvant molecule (data not shown). The active fractions from all four columns had unique

ultraviolet peak absorptions at 235 and 292 nm (Fig. 1a, c–e), which disappeared at low pH (data not shown).

The LMW material from the fourth HPLC column (C18) appeared to be homogeneous. It was subjected to trimethylsilylation (TMS) derivatization and analysed by gas chromatography (GC) coupled with electron impact mass spectrometry (EI-MS; Fig. 2a). A single prominent GC peak was detected at 9.88 min (Fig. 2a). The EI mass spectrum of this peak indicated a compound with an M_r of 456 containing two or more TMS groups. A search of the National Institute of Standards and Technology (NIST) database of EI mass spectra returned a very high probability match with tetra-TMS uric acid (Fig. 2b).

When this same fraction was analysed by direct infusion negative ion electrospray ionization tandem mass spectrometry (ESI-MS), an intense ion at m/z 167 was observed, consistent with the $[M-H]^-$ ion from uric acid. Isolation and subsequent collisional fragmentation of this ion produced an MS^2 product ion spectrum identical to that obtained from a uric acid standard. Similarly, MS^3 product ion spectra from the major MS^2 ion (m/z 124) matched that from uric acid (Fig. 2c–e and Supplementary Fig. 1). The pH-dependent ultraviolet absorption peaks of the LMW fraction at 235 and 292 nm matched well the expected pattern for the enol tautomer of uric acid (Fig. 2f). To exclude the possibility of an enantiomeric or stereo-isomeric structure, the purified LMW molecule was incubated with uricase, a highly specific enzyme that breaks down uric acid to allantoin. Uricase rapidly destroyed the LMW molecule (Fig. 2g). Together, these data show that the main constituent of the highly purified LMW fraction is uric acid.

To evaluate further whether uric acid was responsible for the biological activity seen *in vivo*, we purchased highly purified uric acid and tested it for adjuvant activity. This material enhanced CTL priming to a similar extent to that achieved by the purified LMW fraction (Fig. 3a, b). Comparable results were obtained when C57BL/6 mice were immunized with pure uric acid and particulate ovalbumin antigen (Supplementary Fig. 2). Treatment of the LMW

fraction with uricase significantly reduced its adjuvant activity, providing further proof that uric acid was the active component in this fraction (Fig. 3d, e). Both the purified LMW fraction and the commercial uric acid had adjuvant activity in lipopolysaccharide (LPS)-nonresponsive mice lacking the Toll-like receptor 4 (TLR4; Fig. 3g, h). Thus, LPS is not responsible for the adjuvant activity in our preparations. The finding that uricase inactivated the LMW fraction also argues that LPS or other microbial contaminants do not account for the adjuvant activity (Fig. 3). Because both commercially obtained uric acid and uric acid purified to homogeneity from liver have adjuvant activity and uricase destroys this activity, we conclude that uric acid is one of the endogenous adjuvants in cells.

We have previously reported that endogenous adjuvant activity markedly increases in cells when they are injured in ways that cause them to undergo apoptosis¹⁰. Consistent with this observation, we found that uric acid in EL4 cells increased markedly after treatment with heat shock, cycloheximide and emetine (Fig. 3j), which we have previously shown to increase adjuvant activity. We also observed about a fourfold increase in uric acid when 3T3 cells were treated with ultraviolet radiation (Fig. 3j).

These findings provide insight into the mechanism underlying the increase in endogenous adjuvant in injured cells. Uric acid is produced during the catabolism of purines and is the end product of this process in uricotelic mammals. Thus, as the injured cells rapidly degrade their RNA and DNA, the liberated purines will be converted into uric acid, leading to its accumulation. The production of uric acid does not require protein synthesis, which enables dying cells to produce more danger signal and explains why this increase is not blocked by protein synthesis inhibitors¹⁰.

The amount of uric acid needed to provide an adjuvant effect when injected *in vivo* was about 10 μ g in 100 μ l (Fig. 3k). The active fractions from HPLC purification that were injected into mice all contained more than this amount of uric acid ($>170 \mu$ g ml⁻¹). Ten micrograms of uric acid was produced by about 3×10^6 3T3 cells

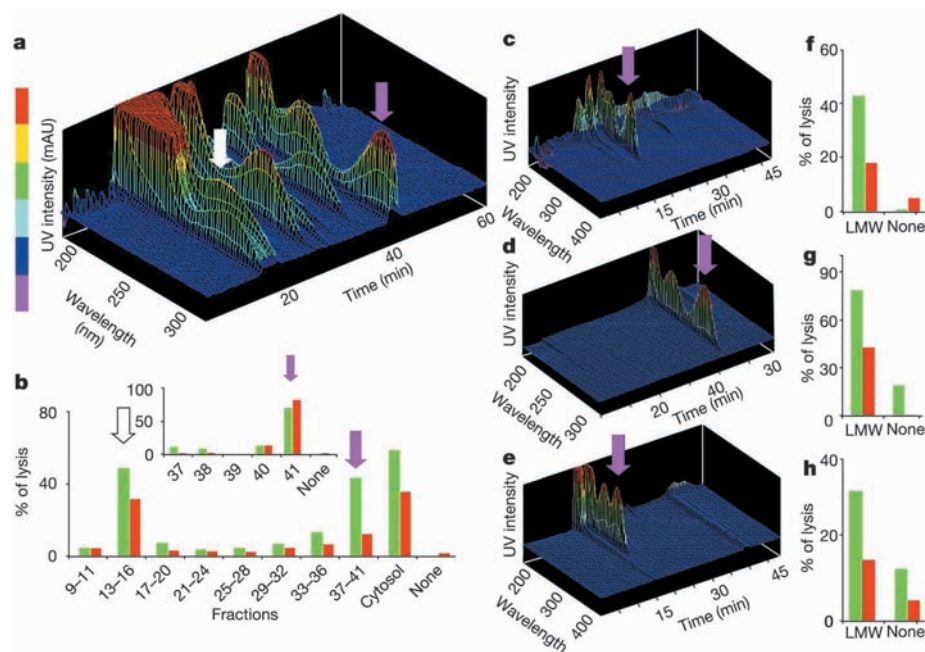


Figure 1 Purification of the LMW endogenous adjuvant from cytosol. **a**, Ultraviolet spectrum of 3T3 cytosolic fractions from a GF250 sizing column. **b**, Anti-HIV-gp120 specific CTL activity from BALB/c mice injected with HIV gp120 antigen beads alone (none), or admixed with either unfractionated cytosol or pooled fractions from **a**. Green and red bars indicate the percentage of specific lysis at effector : target ratios of 100:1

and 33:1, respectively. Inset, individual analysis of LMW fractions 37–41. **c–e**, Ultraviolet spectra of the sequential purification of LMW fractions on anion exchange (**c**), sizing (**d**) and reverse phase (**e**) columns. **f–h**, CTL activity from mice injected with HIV gp120 and the active LMW fractions from **c–e**, respectively. In **a–e**, the active LMW and HMW fractions are indicated by pink and white arrows, respectively.

treated with ultraviolet radiation (Fig. 3). Such numbers of cells could easily die in tumours, and during viral infections and other pathological situations. In fact, it is well known that large amounts of uric acid can be produced from tissue injury *in vivo*¹².

If uric acid is a physiologically important danger signal that promotes immunity, then its elimination *in vivo* should result in reduced T-cell responses to antigen. To test this hypothesis, 3T3 cells were treated with allopurinol (a uric acid analogue that reduces uric acid production) and then injured by ultraviolet irradiation and freezing. These cells were then injected, along with gp120 beads, into mice treated with allopurinol and uricase (to destroy any uric acid that was released *in vivo*). Allopurinol and uricase treatment of mice markedly reduced plasma uric acid concentrations (Supplementary Fig. 3). The priming of gp120-specific CTLs by 10^4 injured cells was substantially reduced by eliminating uric acid (Fig. 4a). Cultures of CTLs from control mice contained 11.8 lytic units as compared with 2.0 lytic units in CTLs from mice treated with allopurinol plus uricase (83% inhibition, Fig. 4b; representative of several experiments). CTLs were still generated in treated mice that were injected with larger numbers of injured cells (10^5). This may be due to the presence of other endogenous adjuvants or the incomplete elimination of uric acid. In the mice treated with allopurinol plus uricase, however, 10^5 injured 3T3 cells stimulated an anti-gp120-specific CTL response equivalent to that induced by 10^4 injured 3T3 cells in control mice (Fig. 4a). In other words, elimination of uric acid reduced the priming of CTLs by about 90%.

To exclude the possibility that the allopurinol and uricase treatment was nonspecifically immunosuppressive, we immunized both treated and control mice with activated bone-marrow-derived

dendritic cells that were pulsed with a synthetic peptide corresponding to the gp120 epitope. As expected, the peptide-pulsed dendritic cells primed CTL responses without any added adjuvant (Fig. 4c). Notably, the generation of anti-gp120 CTLs was not reduced in the treated mice even under limiting conditions (Fig. 4c). The increased response seen in the treated mice was not observed in other experiments. Therefore, the elimination of uric acid *in vivo* selectively reduces T-cell responses that are stimulated by injured cells. We conclude that uric acid is one of the chief mediators produced by injured cells that signals danger and promotes immune responses.

Adjuvants are thought to work, at least in part, by stimulating dendritic cells to mature and to increase their expression of costimulatory molecules^{1,5,6}. We found that when uric acid was added to cultures of primary bone-marrow-derived dendritic cells (>95% CD11C⁺), it stimulated them to increase rapidly (within 6 h) their expression of CD86 and, to a lesser extent, CD80 (Fig. 5a). This stimulation was stronger than that produced by LPS (Fig. 5b). This effect was also seen with dendritic cells from TLR4-null mice and therefore was not due to contaminating LPS (Fig. 5b). Although uric acid stimulated dendritic cells to increase their expression of costimulatory molecules, it did not affect their rate of phagocytosis of antigen particles (Supplementary Fig. 4).

We noted that the concentrations of uric acid that stimulated dendritic cells were also those in which crystals precipitated (Fig. 5c and Supplementary Fig. 5). We also found that preformed monosodium urate (MSU) crystals were highly stimulatory, whereas soluble uric acid was not (Fig. 5b). It is therefore likely that MSU crystals are the biologically active form *in vitro*. It is also possible that crystals are the active form *in vivo*, because uric acid is reported

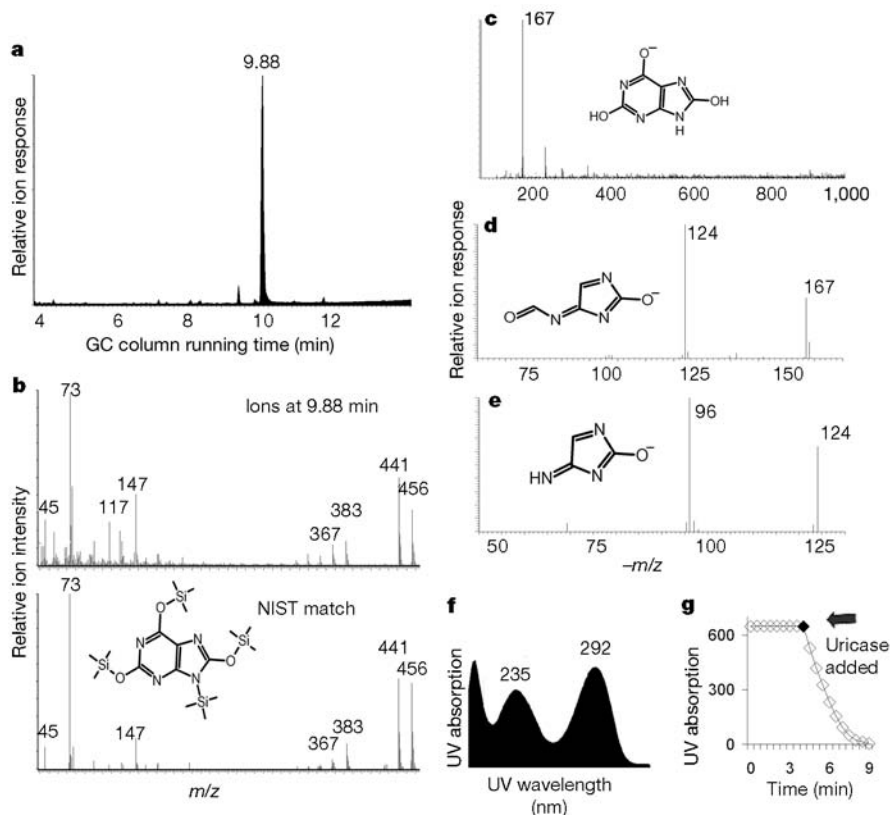


Figure 2 Molecular identification of the LMW endogenous adjuvant. **a**, Total ion current plot from GC-ESI-MS analysis of a TMS-derivatized LMW active fraction purified as in Fig. 1e. **b**, Top, full ESI-MS of the 9.88-min fraction in **a**. Bottom, matching mass spectrum from the NIST database and the structure of tetra-TMS uric acid (inset). **c**, Infusion negative ion ESI-MS spectrum of the same fraction. **d**, **e**, MS² and MS³ product ion

spectra of the m/z 167 and m/z 124 ions from **c** and **d**, respectively. Insets, proposed product ion structures. Uric acid yielded the identical initial ion and subsequent fragments (Supplementary Fig. 1). **f**, Ultraviolet spectrum of a uric acid standard at neutral pH. **g**, Uricase degradation of the highly purified LMW fraction assayed by the reduction of ultraviolet absorption at 292 nm.

to be saturating and to precipitate *in vivo* at the concentrations that we injected ($>70 \mu\text{g ml}^{-1}$)¹³. Although the concentration of uric acid varied from 170 to $1,600 \mu\text{g ml}^{-1}$ in our active HPLC fractions, it remained in solution because it is more soluble in buffers than in body fluids (refs 14–18 and Fig. 1). In addition, the concentration of uric acid in our preparations of liver cytosol is typically about 4 mg ml^{-1} , which when released *in vivo* should locally produce a highly supersaturated condition. Preformed MSU crystals also had adjuvant activity when injected *in vivo* (Fig. 5d). Thus, these data suggest that a chemical phase transition could be the key event that transforms a normal autologous component into a danger signal.

We next investigated whether any crystalline or particulate material would stimulate dendritic cells. Particles of aluminium hydroxide (alum, a weak adjuvant for antibody responses; Fig. 6a) or polystyrene (data not shown) did not stimulate dendritic cells *in vitro* to increase expression of CD86. In marked contrast to MSU crystals (Figs 6a and 5a–c), crystals of allopurinol, a molecule that is structurally very similar to uric acid, or of basic calcium phosphate

(BCP) failed to activate dendritic cells (Fig. 6a). Similarly, the co-injection of BCP crystals, allopurinol crystals or alum *in vivo* all failed to augment T-cell responses to gp120, whereas MSU crystals provided a strong adjuvant effect (Fig. 6b). These results show that MSU crystals have specific ability to activate dendritic cells and to augment the priming of T cells; therefore, these biological effects are unlikely to be due to a simple mechanical stimulation of dendritic cells by particles.

In hyperuricemia, MSU crystals can precipitate in joints, where they cause gout, and/or in other tissues causing inflammation^{13,19}. In these situations, the crystals stimulate monocytes, macrophages and epithelial cells to produce inflammatory mediators^{20–22}. It seems likely that dendritic cells are stimulated in a similar fashion and that the activation of several of these cells of the innate immune system contributes to the adjuvant effect of uric acid. Our findings also raise the possibility that dendritic cells are involved in the pathogenesis of gouty inflammation.

Our findings have several possible implications for health and disease. Up until now, MSU crystals were viewed solely as pathogenic and the biological response to them as pathological. Our findings suggest that the formation of these crystals and the ensuing host response could have an important role in immune surveillance and the generation of adaptive immunity. In addition, it is well known that *in vivo* dying tissue initially provokes a strong inflammatory response but loses this phlogistic quality as it is depleted of its unidentified proinflammatory mediators²³. Our data raise the possibility that uric acid is one of the mediators that leads to the inflammatory response to injured and dying tissues. Thus MSU might also be involved in inflammation to injury.

In terms of immune surveillance, our data suggest a model in which dying cells release uric acid together with their antigens. Dendritic cells ingest the cellular debris and thereby acquire antigens from the dying cells. At the same time they receive a 'danger signal' from the uric acid (and maybe other endogenous adjuvants) and are stimulated to mature and to become immunostimulatory. Thus, uric acid alerts the immune system to cell death, which is a

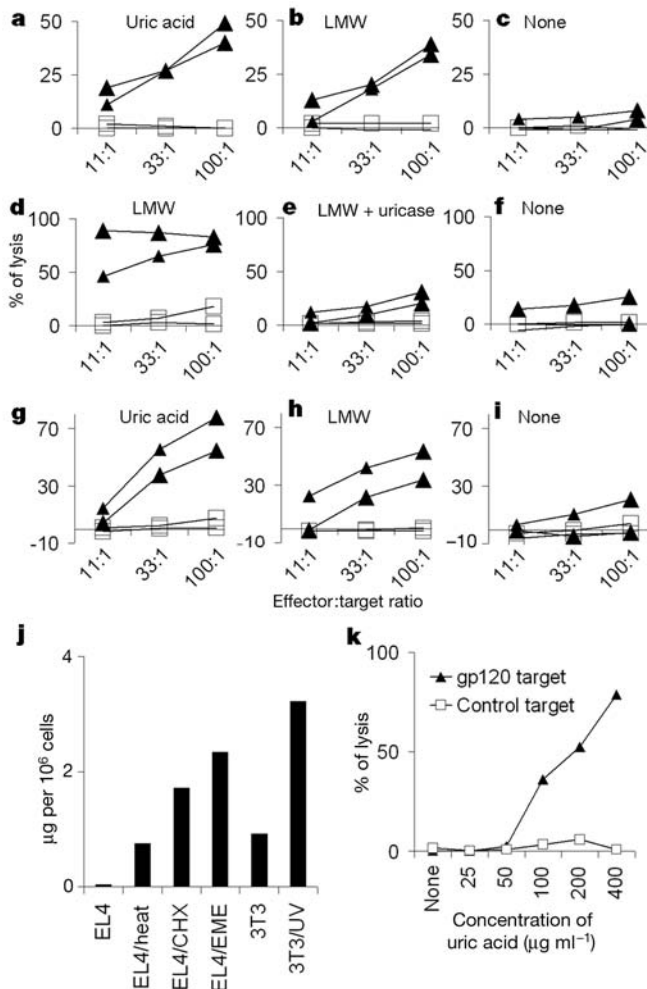


Figure 3 Uric acid has adjuvant activity *in vivo* and its concentrations increase in injured cells. **a–c**, CTLs from BALB/c mice immunized with uric acid (50 μg) (**a**), the LMW fraction (purified as in Fig. 1e) (**b**) or PBS (**c**) mixed with 5 μg of gp120-latex beads were assayed on gp120⁺ (filled triangles) or control (open squares) targets; data points represent individual mice. **d–f**, As **a–c**, except that untreated (**d**) or uricase-treated (**e**) LMW fractions were injected. **g–i**, As **a–c**, except that C3H/HeJ (*Tlr4*^{−/−}) mice were used. **j**, Uric acid content in EL4 or 3T3 cells after no treatment, or treatment with heat shock (heat), cycloheximine (CHX), emetine (EME) or ultraviolet irradiation (UV). **k**, As **a**, except that the amount of uric acid co-injected was varied. Values are the mean lysis at an effector:target ratio of 100:1 for two mice.

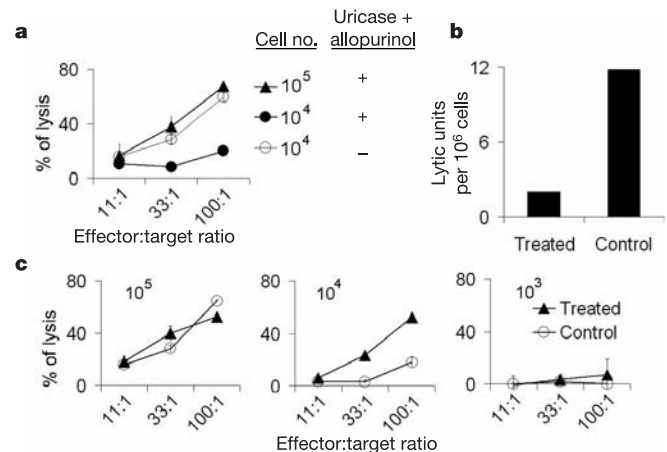


Figure 4 Eliminating uric acid *in vivo* inhibits adjuvant activity from injured cells. **a**, Mice treated with uricase and allopurinol were immunized with 10^4 (filled circles) or 10^5 (filled triangles) injured 3T3 cells (Methods) treated with uricase and allopurinol plus gp120-latex beads (2 μg), and assayed as in Fig. 1b. Control mice (open circles) were assayed similarly except that uricase and allopurinol were omitted from all steps. Data are the mean $\pm$ range of lysis for two mice. **b**, Analysis of lytic units from the cells in **a** (treated versus untreated mice receiving 10^4 cells). **c**, 10^5 (left), 10^4 (middle) or 10^3 (right) gp120 peptide-pulsed dendritic cells were injected s.c. into mice that had been treated with (filled triangles) or without (open circles) uricase and allopurinol, and assayed as in Fig. 1b (Methods).

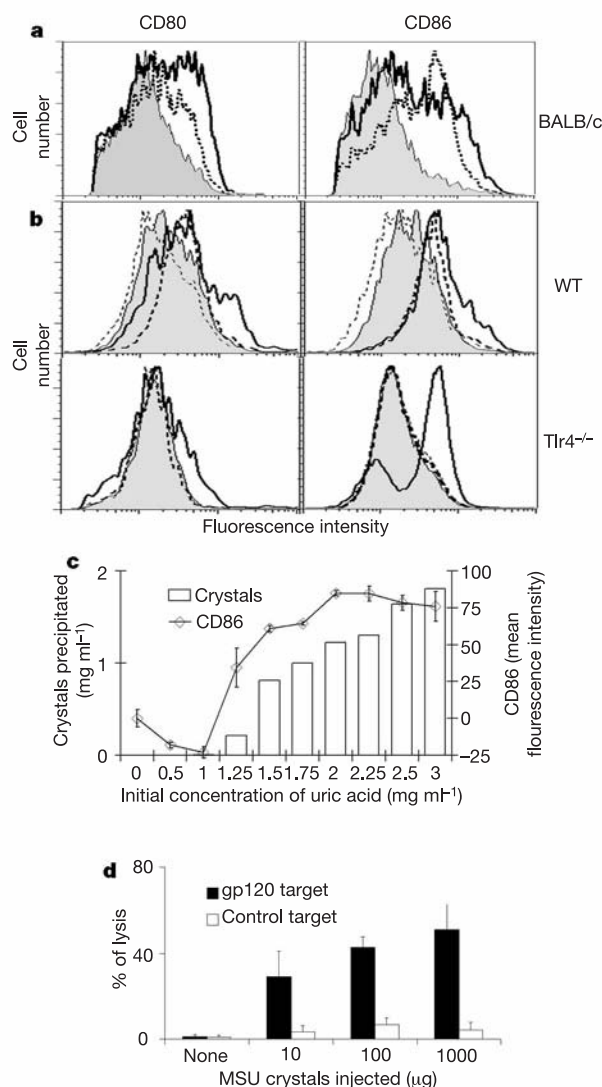


Figure 5 MSU crystals rapidly activate dendritic cells. **a**, Expression of CD80 and CD86 on BALB/c bone-marrow-derived dendritic cells that were either untreated (shaded area) or incubated with 20 μ l of MSU crystals for 6 h (broken lines) or 24 h (unbroken lines). **b**, As **a**, except that dendritic cells from wild-type C57BL/6 (WT) or C3H/HeJ (*Tlr4*^{-/-}) mice were untreated (shaded areas), incubated with soluble uric acid (70 μ g ml⁻¹, thin broken lines), 1 μ g ml⁻¹ LPS (thick broken lines) or 20 μ l of MSU crystals per well (thick unbroken lines) for 6 h. **c**, Expression of CD86 on dendritic cells (diamonds) versus the quantity of MSU crystals forming in cultures (bars) incubated for 24 h with the indicated concentrations of uric acid. **d**, As Fig. 3a, except that preformed MSU crystals were injected into mice, and data are the mean $\pm$ range of lysis at an effector : target ratio of 100:1 of gp120-transfected (filled bars) or control transfected (open bars) targets for two mice.

hallmark of potential danger to the host. If the dying cell contains antigens to which the host is not tolerant, then an immune response will be stimulated. This therefore provides a mechanism by which the immune system could generate responses against tumours and viral infections that were thought to lack adjuvants.

Our *in vitro* data indicate that the immunostimulatory form of uric acid may be MSU crystals. Uric acid is relatively insoluble in biological fluids (70 μ g ml⁻¹) and mammals have relatively high constitutive concentrations of uric acid in the blood (40–60 μ g ml⁻¹). Because cytosol contains high concentrations of uric acid (for example, 4 mg ml⁻¹, which increases even more when injured cells degrade their RNA and DNA), the local environment

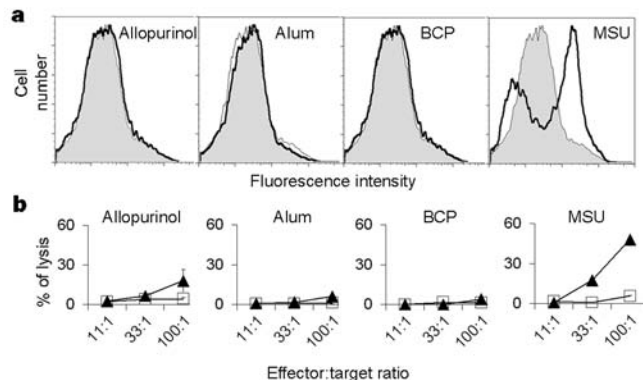


Figure 6 Specificity of dendritic cell activation by MSU crystals. **a**, As Fig. 5a except that BALB/c bone-marrow-derived dendritic cells were incubated for 6 h in 2 ml of media containing 0.5 mg ml⁻¹ of crystals of allopurinol, BCP or MSU, or alum (Methods). Unbroken line indicates CD86 expression, the shaded curve indicates background (untreated dendritic cells). **b**, BALB/c mice were immunized with 500 μ g of the indicated crystals or alum admixed with gp120-latex beads (2 μ g) and assayed as in Fig. 5d. Values are the mean $\pm$ range of lysis of targets by CTLs from two mice.

around the dying cells should become supersaturated with uric acid released when cell lyse. This would favour the formation of MSU. The MSU crystals would also stimulate monocytes and epithelial cells to produce cytokines, which may further potentiate immune responses¹³. In the same way, the danger signal provided by uric acid and other endogenous adjuvants might underlie the initiation of autoimmunity in genetically susceptible individuals. In this context, individuals affected with some autoimmune diseases have been reported to have a higher incidence of hyperuricaemia than controls^{19,24}.

There also remains a need for the development of more and better adjuvants for use in humans. Alum, the only adjuvant approved for human use, is relatively weak, does not generally elicit CD8 immunity, and may preferentially induce responses biased toward T-helper type 2 cells¹. Uric acid represents a previously unknown class of adjuvant that is distinct from those of microbial origin. It stimulates strong CD8⁺ T-cell immunity. Our findings therefore indicate that uric acid may be useful as a type of adjuvant for vaccines.

In summary, we have identified uric acid as one of the principal endogenous immunological danger signals. It is constitutively present in cells and its concentration increases when cells are injured. When released from dying cells, it stimulates dendritic cells to mature and augments the priming of CD8⁺ T-cell responses to cross-presented antigen. □

Methods

Mice, cells and reagents

All biochemical reagents were from Sigma. The Amplex Red kit for detecting uric acid was from Molecular Probes. MSU crystals were prepared by incubating supersaturated uric acid solutions (4–5 mg ml⁻¹) in 0.1 M borate (pH 8.5) at room temperature (26 °C) for >48 h, followed by washing with alcohol and acetone. Preparations of other crystals are described in Supplementary Information. We prepared antigen beads as described^{10,25}. Peptides have been described¹⁰ and were a gift from T. Vedvick (Corixa). Mice, cells and all other reagents have been described^{10,11}.

Purification and treatment of LMW adjuvant

We produced ultraviolet-treated 3T3 cells and liver cytosol as described¹⁰. Chromatographic separations were done with a System Gold 125 solvent module linked to a System Gold 168 diode array spectrum analyser running Gold Nouveau software v1.0 (Beckman). Cytosolic fractions were purified sequentially on Zorbax GF-250 (Agilent) or Superdex 75 HR 10/30 (Pharmacia), Mono Q HR 5/5 (Pharmacia), Superdex 200 HR 10/30 (Pharmacia) and semi-preparative C18 (Vydac) columns. HPLC procedures and buffers are described in the Supplementary Information. Where indicated, the LMW C18 fraction (containing 1/40 liver equivalent) was incubated in 500 μ l of 0.1 M borate buffer (pH 8.5) at room temperature (26 °C), 0.01 unit of uricase was added, and the ultraviolet

absorption at 292 nm was monitored. In some experiments, the LMW C18 fraction (containing 1/10 liver equivalent) was treated with 0.1 units of uricase at room temperature (26 °C) before being used for injection *in vivo*.

Immunization and CTL assays

Immunizations and CTL assays were done as described¹⁰ except that the indicated adjuvants were used. Column fractions from 1–10% liver cytosol were injected without manipulation (GF250 column) or after drying and reconstitution (other columns). Pooled fractions without adjuvant activity were usually used as negative controls. For *in vitro* stimulation, splenocytes from immunized mice were stimulated with 10^{-8} M of RGPGRFVFTI or SIINFEKL peptide directly, and assayed as described¹⁰.

In some experiments, mice were injected with uricase (10 µg per mouse) and allopurinol (800 µg per mouse) in the peritoneum daily for 3 d, and 3T3 cells were cultured in 50 µM allopurinol. The cells were irradiated with ultraviolet as described¹⁰, frozen for 20 min, and varying numbers were then admixed with gp120–latex beads and injected subcutaneously (s.c.) together with allopurinol (20 µg) and uricase (0.4 µg) into the pretreated mice. Alternatively, mice treated with allopurinol plus uricase were immunized s.c. with varying numbers of bone-marrow-derived dendritic cells that had been pulsed with HIV gp120 peptide RGPGRFVFTI (1 µg ml⁻¹ for 1 h). Controls were treated identically except that uricase and allopurinol were omitted from all steps.

Mass spectrometry

GC-MS and ESI-MS were done with a Quattro-II triple quadrupole (Waters) and a LCQ quadrupole ion trap (Finnigan) mass spectrometer, respectively. Technical parameters, as well as the method of TMS derivatization, are given in the Supplementary Information.

Uric acid measurement

EL4 cells (10^8) were resuspended in 5 ml of culture media and incubated at 37 °C for 5 h after being treated with emetine¹⁰ or cycloheximide for 1 h or heat-shocked at 45 °C for 20 min. 3T3 cells were either untreated or exposed to ultraviolet light for 5 min and incubated for 5 h. The cells were directly suspended in culture media (without washing) and disrupted by nitrogen cavitation followed by centrifugation. We determined the concentration of uric acid in cytosol, HPLC fractions and mouse plasma by either HPLC or a uric acid kit.

Dendritic cell culture and analysis

We cultured bone-marrow-derived dendritic cells as described²⁶. After removing floating cells, cultures were stimulated as indicated for 6 or 24 h. Bound crystals were detached by incubation with either 0.5% heparin or 0.5 µM polyvinyl sulphate (100K) in PBS for 10 min. The cells were then collected by scraping and extensively washed. Staining and flow cytometry have been described¹¹. MSU crystals were collected from parallel sets of wells without dendritic cells, washed with absolute alcohol and dissolved in 0.1 N NaOH. The amount of uric acid was determined by measuring ultraviolet absorption at 292 nm against a set of standard uric acid and NaOH solutions.

Received 9 July; accepted 18 August 2003; doi:10.1038/nature01991.

Published online 7 September 2003.

- Hunter, R. L. Overview of vaccine adjuvants: present and future. *Vaccine* **20** (suppl. 3), S7–S12 (2002).
- Janeway, C. A. Jr Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **54**, 1–13 (1989).
- Janeway, C. A. Jr The immune system evolved to discriminate infectious nonself from noninfectious self. *Immunol. Today* **13**, 11–16 (1992).
- Matzinger, P. Tolerance, danger, and the extended family. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
- Cox, J. C. & Coulter, A. R. Adjuvants—a classification and review of their modes of action. *Vaccine* **15**, 248–256 (1997).
- Schijns, V. E. Immunological concepts of vaccine adjuvant activity. *Curr. Opin. Immunol.* **12**, 456–463 (2000).
- Medzhitov, R. & Janeway, C. Jr Innate immune recognition: mechanisms and pathways. *Immunol. Rev.* **173**, 89–97 (2000).
- Albert, M. L., Sauter, B. & Bhardwaj, N. Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs. *Nature* **392**, 86–89 (1998).
- Gallucci, S., Lolkema, M. & Matzinger, P. Natural adjuvants: endogenous activators of dendritic cells. *Nature Med.* **5**, 1249–1255 (1999).
- Shi, Y., Zheng, W. & Rock, K. L. Cell injury releases endogenous adjuvants that stimulate cytotoxic T cell responses. *Proc. Natl Acad. Sci. USA* **97**, 14590–14595 (2000).
- Shi, Y. & Rock, K. L. Cell death releases endogenous adjuvants that selectively enhance immune surveillance of particulate antigens. *Eur. J. Immunol.* **32**, 155–162 (2002).
- Jeha, S. Tumor lysis syndrome. *Semin. Hematol.* **38**, 4–8 (2001).
- Smyth, C. J. & Holers, V. M. *Gout, Hyperuricemia, and Other Crystal-Associated Arthropathies* (Marcel Dekker, New York, 1998).
- Kippen, I., Klinenberg, J. R., Weinberger, A. & Wilcox, W. R. Factors affecting urate solubility *in vitro*. *Ann. Rheum. Dis.* **33**, 313–317 (1974).
- Iwata, H., Nishio, S., Yokoyama, M., Matsumoto, A. & Takeuchi, M. Solubility of uric acid and supersaturation of monosodium urate: why is uric acid so highly soluble in urine? *J. Urol.* **142**, 1095–1098 (1989).
- Fiddis, R. W., Vlachos, N. & Calvert, P. D. Studies of urate crystallisation in relation to gout. *Ann. Rheum. Dis.* **42** (suppl. 1), 12–15 (1983).
- Tak, H. K., Cooper, S. M. & Wilcox, W. R. Studies on the nucleation of monosodium urate at 37 °C. *Arthritis Rheum.* **23**, 574–580 (1980).
- Loeb, J. N. The influence of temperature on the solubility of monosodium urate. *Arthritis Rheum.* **15**, 189–192 (1972).
- Talbott, J. H. & Yu, T.-F. *Gout and Uric Acid Metabolism* (Stratton Intercontinental Medical Book Corp., New York, 1976).

- Yagnik, D. R. *et al.* Noninflammatory phagocytosis of monosodium urate monohydrate crystals by mouse macrophages. Implications for the control of joint inflammation in gout. *Arthritis Rheum.* **43**, 1779–1789 (2000).
- Landis, R. C. *et al.* Safe disposal of inflammatory monosodium urate monohydrate crystals by differentiated macrophages. *Arthritis Rheum.* **46**, 3026–3033 (2002).
- Koka, R. M., Huang, E. & Lieske, J. C. Adhesion of uric acid crystals to the surface of renal epithelial cells. *Am. J. Physiol. Renal Physiol.* **278**, F989–F998 (2000).
- Majno, G. & Joris, I. *Cells, Tissues, and Disease: Principles of General Pathology* (Blackwell Science, Cambridge, MA, 1996).
- Bosmansky, K. & Trnavsky, K. Serum uric acid levels in disorders of the rheumatic type. *Z. Rheumatol.* **43**, 59–62 (1984).
- Vidard, L. *et al.* Analysis of MHC class II presentation of particulate antigens of B lymphocytes. *J. Immunol.* **156**, 2809–2818 (1996).
- Coligan, J. E., Kruisbeek, A. M., Margulies, D. H., Shevach, E. M. & Strober, W. *Current Protocols in Immunology* (Jon Wiley & Sons, Hoboken, NJ, 1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Sodrosky, J. Berzofsky, N. Sato, Y. Takeda and T. Vedvick for reagents; B. Zhang, P. Furcinitti and B. Kobertz for advice and technical assistance; and L. Stern and I. York for critically reviewing the manuscript. The work was supported by NIH grants to K.L.R. Y.S. was also supported by a grant from the Pittsfield TB association.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.L.R. (kenneth.rock@umassmed.edu).

Differential regulation of EIN3 stability by glucose and ethylene signalling in plants

Shuichi Yanagisawa^{1,2}, Sang-Dong Yoo² & Jen Sheen²

¹Research Institute for Bioresources, Okayama University, Chuo 2-20-1, Kurashiki 710-0046, Japan

²Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA

Glucose is a global regulator of growth and metabolism that is evolutionarily conserved from unicellular microorganisms to multicellular animals and plants¹. In photosynthetic plants, glucose shows hormone-like activities and modulates many essential processes, including embryogenesis, germination, seedling development, vegetative growth, reproduction and senescence^{2,3}. Genetic and phenotypic analyses of *Arabidopsis* mutants with glucose-insensitive (*gin*) and glucose-oversensitive (*glo*) phenotypes have identified an unexpected antagonistic interaction between glucose and the plant stress hormone ethylene. The ethylene-insensitive *etr1* and *ein2* mutants have *glo* phenotypes, whereas the constitutive ethylene signalling mutant *ctr1* is allelic to *gin4* (refs 4, 5). The precise molecular mechanisms underlying the complex signalling network that governs plant growth and development in response to nutrients and plant hormones are mostly unknown. Here we show that glucose enhances the degradation of ETHYLENE-INSENSITIVE3 (EIN3), a key transcriptional regulator in ethylene signalling^{6,7}, through the plant glucose sensor hexokinase⁸. Ethylene, by contrast, enhances the stability of EIN3. The *ein3* mutant has a *glo* phenotype, and overexpression of EIN3 in transgenic *Arabidopsis* decreases glucose sensitivity.

To test the hypothesis that glucose modulates ethylene signalling through the transcription factor EIN3, which acts downstream of ETR1 (ref. 9) and EIN2 (ref. 10), we tested the activity of EIN3 protein tagged with the Myc epitope in maize mesophyll proto-

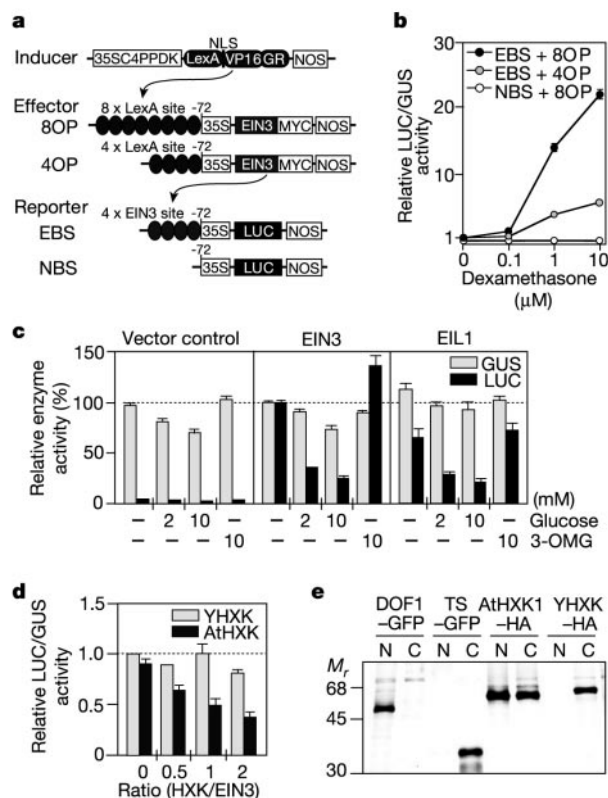


Figure 1 Functional analysis of EIN3 in maize protoplasts. **a**, The inducible expression system. 8OP and 4OP, LexA operator binding sites; LexA, LexA DNA-binding domain; NLS, nuclear localization signal; VP16, transcriptional activation domain; GR, glucocorticoid receptor. EBS indicates with, and NBS indicates without, the EIN3-binding site. **b**, Dosage-dependent transcriptional activation by EIN3. **c**, Glucose represses EIN3 and EIL1 activity. *EBS-LUC* (LUC) and *UBQ-GUS* (GUS) activities are shown. **d**, AtHXK1-dependent repression. **e**, Nuclear association of AtHXK1. Equal amounts of the nuclear (N) and cytosolic (C) fractions prepared from the same protoplast samples were analysed. TS, plastid transit peptide.

plasts, a well-established glucose-responsive plant-cell system¹¹. To control EIN3 expression and to ensure specificity, we first used a glucocorticoid-receptor-inducible gene expression system based on the bacterial LexA repressor and its cognate DNA-binding site (ref. 12 and Fig. 1a). A luciferase (LUC) reporter gene controlled by four copies of the EIN3-binding sites (*EBS-LUC*)⁷ and the 35S basal promoter¹³ was used to quantify the activity of EIN3. In all transient expression assays, we used another reporter gene encoding β -glucuronidase (GUS) under the control of a constitutive ubiquitin promoter as an internal control^{14,15}. We found that EIN3-dependent activation of *EBS-LUC* was markedly induced by dexamethasone in a concentration-dependent manner (Fig. 1b), showing that EIN3 is sufficient to activate transcription in maize protoplasts through a specific DNA-binding site that was previously defined *in vitro* for *Arabidopsis* EIN3 (ref. 7) and a tobacco EIN3 homologue TEIL¹⁶.

To examine whether glucose influences EIN3 activity in maize protoplasts, we measured the transcriptional activity of EIN3 in the presence of glucose and a non-signalling glucose analogue, 3-O-methyl-glucose (3-OMG)¹¹, as an osmotic control. Expression of EIN3 from a constitutive CaMV 35S promoter strongly induced *EBS-LUC*, and only glucose at physiological concentrations (2 and 10 mM) significantly suppressed this *EBS-LUC* activity (Fig. 1c). Using the same cell system, we further tested the transcriptional activity and glucose sensitivity of an *Arabidopsis* EIN3-like protein (EIL1) that is known to rescue the *ein3* mutant phenotype⁶. EIL1 activated the expression of *EBS-LUC*, and this activation was repressed by glucose but not by 3-OMG (Fig. 1c). Thus, *Arabidopsis* EIN3 and EIL1 have similar DNA-binding and transcriptional activation functions and similar sensitivity to glucose.

We previously showed that *Arabidopsis thaliana* hexokinase (AtHXK1) acts as a glucose sensor with both metabolic and signalling functions⁸. Increasing the expression of AtHXK1 but not yeast HXK2 (YHXX2) led to stronger repression of EIN3-dependent *EBS-LUC* expression (Fig. 1d). Notably, immunoprecipitation analysis of the haemagglutinin (HA)-tagged HXKs showed that AtHXK1 but not YHXX2 could associate with the nuclear fractions of maize protoplasts (Fig. 1e). Fractionation was verified with two marker proteins tagged with green fluorescent protein (GFP), DOF1-GFP¹⁴ and TP-GFP¹⁷, which localize in the nucleus and chloroplasts, respectively. TP-GFP but not DOF1-GFP was released to the cytoplasmic fraction by 1% Triton X-100 (Fig. 1e).

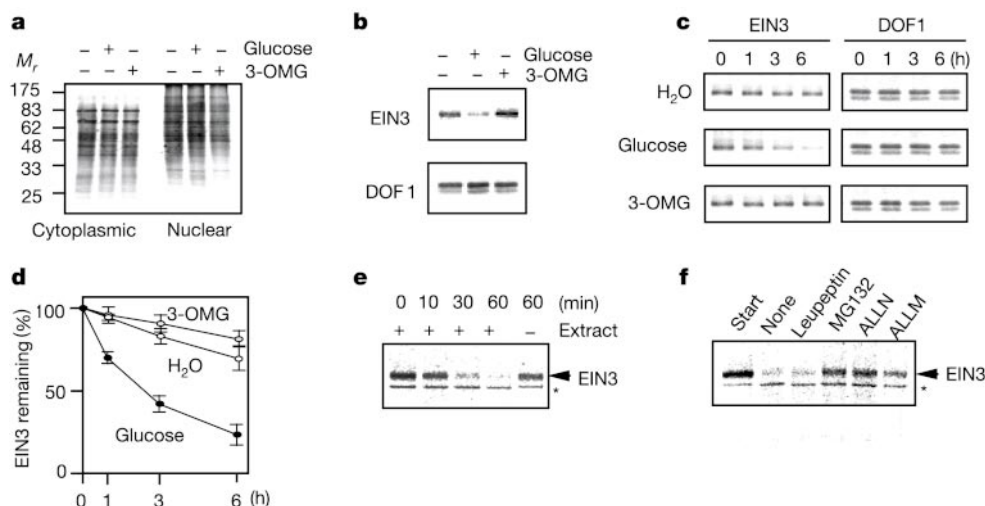


Figure 2 Glucose accelerates EIN3 degradation. **a**, Profiles of labelled proteins. Cytoplasmic fraction from 1.6×10^4 protoplasts or nuclear extract from 2×10^5 nuclei was loaded with a similar quantity of radioactivity. **b**, Accumulation of EIN3-Myc is affected by glucose. **c**, EIN3 degradation is enhanced by glucose *in vivo*. **d**, Plot of the

amounts of EIN3 remaining. **e**, Time course of EIN3 degradation in a cell-free system. **f**, Cell-free degradation of EIN3 is inhibited by proteasome-specific inhibitors. Degradation reactions were allowed to proceed for 1 h with 10 μ M of each inhibitor. Asterisks indicate nonspecific bands.

Although HXK is generally considered to be a soluble enzyme that is involved in glycolysis in the cytoplasm⁸, AtHXK1 associated with the nucleus might directly or indirectly mediate glucose signalling into or out of the nucleus and might regulate EIN3 activity.

As *EIN3* gene expression is not induced by ethylene, it has been suggested that ethylene might activate *EIN3* (ref. 6) or might regulate *EIN3* protein expression¹⁸. To elucidate the molecular mechanism underlying the glucose-mediated repression of *EIN3* activity, we investigated the stability of *EIN3* protein. Constructs expressing *EIN3* tagged with Myc and the unrelated maize transcription factor *DOF1* tagged with HA¹⁴ were co-transfected into protoplasts, and the proteins were labelled with [³⁵S]Met in the presence or absence of 10 mM glucose or 3-OMG. Neither glucose nor 3-OMG affected the labelling patterns of total cytoplasmic and nuclear proteins (Fig. 2a), or the total amount of protein synthesis (data not shown). However, the amount of *EIN3*, but not *DOF1*, was specifically diminished by glucose but not by 3-OMG (Fig. 2b).

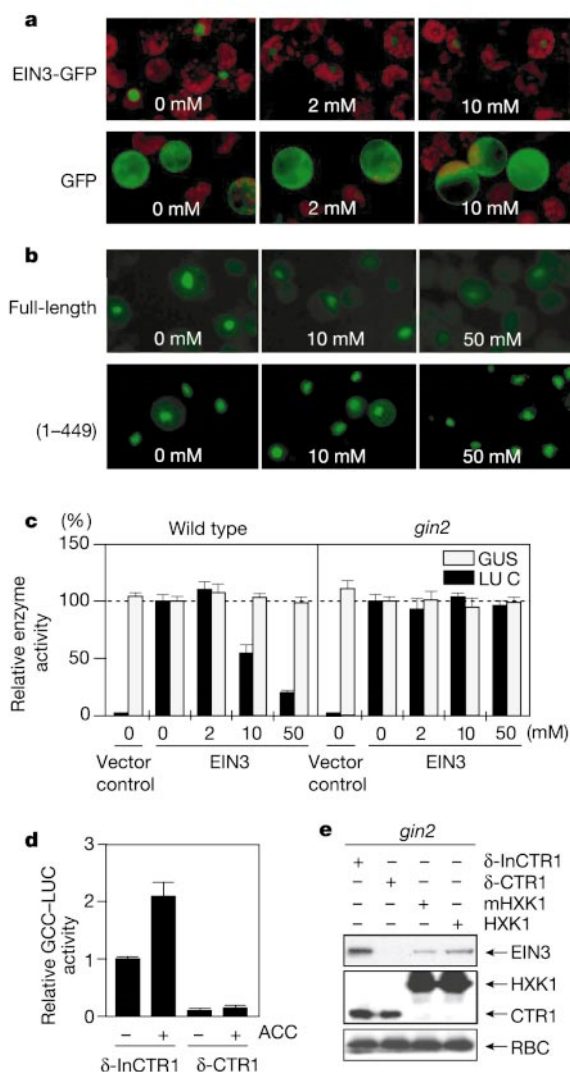


Figure 3 Differential glucose and ethylene signalling. **a**, Nuclear accumulation of *EIN3*-GFP in maize protoplasts. The red autofluorescence is from chlorophyll. The glucose concentration is shown. **b**, Nuclear accumulation of *EIN3*-GFP in *Arabidopsis* protoplasts. The GFP-tagged N terminus of *EIN3* (residues 1–449) is insensitive to glucose. **c**, Glucose repression of *EIN3* activity is abolished in the *gin2* mutant protoplasts. LUC, *EBS-LUC*; GUS, *UBQ10-GUS*. **d**, Ethylene and ACC activation of *GCC-LUC* is abolished by CTR1. δ -CTR1, constitutively active CTR1; δ -InCTR1, kinase mutant of CTR1. **e**, *EIN3* accumulation is reduced by CTR1 and HXK1 in the *gin2* mutant protoplasts. mHXK1, catalytically inactive HXK1; RBC, Rubisco.

Thus, glucose might accelerate the degradation of *EIN3*, thereby reducing the activation of *EBS-LUC*.

To monitor the degradation of *EIN3* directly *in vivo*, we incubated the transfected and labelled protoplasts in the presence of 10 mM glucose or 3-OMG for 1, 3 and 6 h. The amount of *EIN3* protein was determined by immunoprecipitation. It was evident that *EIN3* degradation was much faster in the presence of glucose than in the presence of 3-OMG or in the water control (Fig. 2c, d). To determine the nature of the proteolytic machinery involved in *EIN3* degradation, we established a cell-free degradation assay^{19,20} and tested the ability of different protease inhibitors to stabilize *EIN3*. Purified *EIN3* was completely degraded within 60 min when incubated with maize protoplast cell extract (Fig. 2e). MG132, a potent inhibitor of the 26S proteasome^{19–21}, blocked degradation of *EIN3* (Fig. 2f). Two other proteasome inhibitors, ALLM and ALLN²¹, similarly hampered *EIN3* degradation, whereas the general protease inhibitor leupeptin did not.

To investigate the subcellular localization of *EIN3* in living cells, we fused *EIN3* to jellyfish GFP for easy visualization^{14,15,17}. *EIN3*-GFP accumulated in the nucleus (Fig. 3a), consistent with the location of an *EIN3*-GUS fusion protein⁶. Glucose decreased the accumulation of *EIN3*-GFP in the nucleus (Fig. 3a), similar to its effect on Myc-tagged *EIN3* (Fig. 2b–d). The stability of GFP alone was not affected by glucose (Fig. 3a).

To see whether a similar glucose response might occur in another plant system, we examined the expression of two *EIN3*-GFP proteins in the presence or absence of glucose by using an *Arabidopsis* mesophyll protoplast transient assay¹⁵. As observed in maize protoplasts, GFP-tagged full-length *EIN3* was detected in the nucleus but diminished with increasing concentrations of glucose (Fig. 3b). A GFP-tagged amino-terminal fragment (residues 1–449) of *EIN3* accumulated in the nucleus but was insensitive to glucose (Fig. 3b), suggesting that the carboxy terminus of *EIN3* is involved in the glucose response.

We also showed that the *EIN3*-dependent expression of *EBS-LUC* was repressed by glucose in *Arabidopsis* protoplasts (Fig. 3c). To confirm the requirement of AtHXK1 in the glucose-mediated repression of *EIN3* activity, a transient expression analysis was carried out with protoplasts of the *Arabidopsis* AtHXK1-null mutant *gin2* (refs 3, 8). As shown in maize protoplasts (Fig. 1d), AtHXK1 also controlled *EIN3*-dependent expression of *EBS-LUC* in *Arabidopsis* protoplasts (Fig. 3c). Thus, the regulatory mechanism mediating glucose-dependent degradation of *EIN3* is probably conserved in maize and *Arabidopsis*.

Arabidopsis protoplasts, however, seemed to be less sensitive to glucose because a higher glucose concentration was required to reach the effect observed in maize protoplasts (Figs 1c and 3a–c). Because *Arabidopsis* protoplasts were isolated from 3–4-week-old mature leaves, in which endogenous ethylene signalling is more active than in seedlings, these cells are probably more resistant to glucose than are maize cells isolated from the leaves of 11-day-old seedlings^{11,13}.

To examine ethylene responses in *Arabidopsis* protoplasts, we tested a *LUC* reporter gene controlled by a well-characterized ethylene responsive enhancer *GCC (GCC-LUC)*²². As expected, *GCC-LUC* was active in *Arabidopsis* protoplasts. Addition of the immediate ethylene precursor 1-aminocyclopropane-1-carboxylic acid (ACC) further enhanced its activity, suggesting that these cells are responsive to ethylene (Fig. 3d). To provide further evidence that *GCC-LUC* activity is due to ethylene signalling in protoplasts, we introduced a constitutively active CTR1 variant (ref. 23), which has been identified as a key negative regulator of ethylene signalling²⁴. Active CTR1 abolished *GCC-LUC* activity in *Arabidopsis* protoplasts by preventing the accumulation of *EIN3* in wild-type and *gin2* protoplasts (Fig. 3d, e). Catalytically active or inactive AtHXK1 (ref. 8) was effective in diminishing *EIN3* accumulation in *gin2* protoplasts (Fig. 3e). Thus, *EIN3* is a common target of both the

ethylene and the glucose signalling pathways, which are uncoupled from glucose metabolism.

To investigate the link between the stability of EIN3 and the phenotypes associated with glucose and ethylene signalling *in planta*, we generated transgenic *Arabidopsis* plants expressing Flag-tagged EIN3 under the control of a constitutive 35S promoter¹⁵. The typical ethylene-induced triple response in dark-grown *Arabidopsis* seedlings was examined using ACC^{4,6,18}. The 35S::EIN3-FLAG seedlings showed a much stronger triple response than did wild-type seedlings, indicating that the EIN3-Flag protein is active in mediating ethylene responses. The *ein3* mutant⁶ was insensitive to ACC under the same assay conditions (Fig. 4a). Although EIN3 transcripts from both the endogenous gene and the transgene were present constantly (Fig. 4b), EIN3-Flag was detected only with ACC treatment, suggesting that ethylene enhanced the accumulation and/or stability of EIN3 (Fig. 4c).

To ensure that EIN3-Flag was regulated in the same way as endogenous EIN3, an EIN3-specific antibody was raised and used to detect ACC-dependent EIN3 accumulation (Fig. 4c). 35S::EIN3-FLAG transgenic seedlings were also more resistant to glucose-induced developmental arrest⁴ than were wild type. By contrast, the *ein3* mutant was much more sensitive to glucose than were wild-type seedlings. All plants grew similarly on MS (Murashige-Skoog) plates containing mannitol (2–6%), supporting the idea that EIN3 mediates specific responses to glucose and ethylene but not osmotic stress (Fig. 4d).

To connect these cellular observations to the whole plant, we monitored the degradation of EIN3-Flag in transgenic *Arabidopsis* seedlings in the presence of glucose or ethylene. Seedlings germinated on MS plates containing ACC were incubated in liquid MS medium containing 6% glucose or 6% mannitol and cycloheximide to block *de novo* protein synthesis. Immunoblot analysis showed

that EIN3 degradation proceeded much faster in glucose-treated seedlings than in mannitol- or water-treated seedlings (Fig. 4e). In addition, a proteasome-specific inhibitor, MG132, blocked the glucose-enhanced degradation of EIN3 (Fig. 4f). We also found that the presence of ACC delayed EIN3 degradation (Fig. 4g).

Sugars have been known for decades to influence plant growth and development profoundly, and the most prevailing view has been that sugars simply provide a source of energy and carbon. But extensive genetic analyses of sugar signalling mutants showing *gin* and *glo* phenotypes have identified an unexpectedly intimate connection between the glucose and plant hormone signalling pathways in the model plant *Arabidopsis*^{2–5}. Of these findings, an antagonistic interaction between glucose and the plant stress hormone ethylene has been the least expected⁴. Our results now identify a previously unknown glucose signalling mechanism involving degradation of the EIN3 transcription factor, a well-established key regulator of ethylene signalling^{6,7}. This provides the first molecular link in the plant signalling network that connects seemingly unrelated and distinct signal transduction pathways.

Because the acceleration of EIN3 degradation by glucose was observed in maize and *Arabidopsis*, glucose-dependent protein degradation might be a conserved mechanism for controlling growth and development in plants. Under physiological conditions, endogenous glucose signals probably promote growth by antagonizing stress hormone ethylene signalling and potentiating growth hormone auxin signalling⁸. In both plants and animals, monitoring endogenous glucose concentrations could have a central regulatory role in coordinating nutrient metabolism and hormone synthesis and signalling, which are also influenced by environmental conditions such as light, temperature and stress in plants.

The regulation of protein stability has recently emerged as a key mechanism that controls photomorphogenesis²⁰ and plant hor-

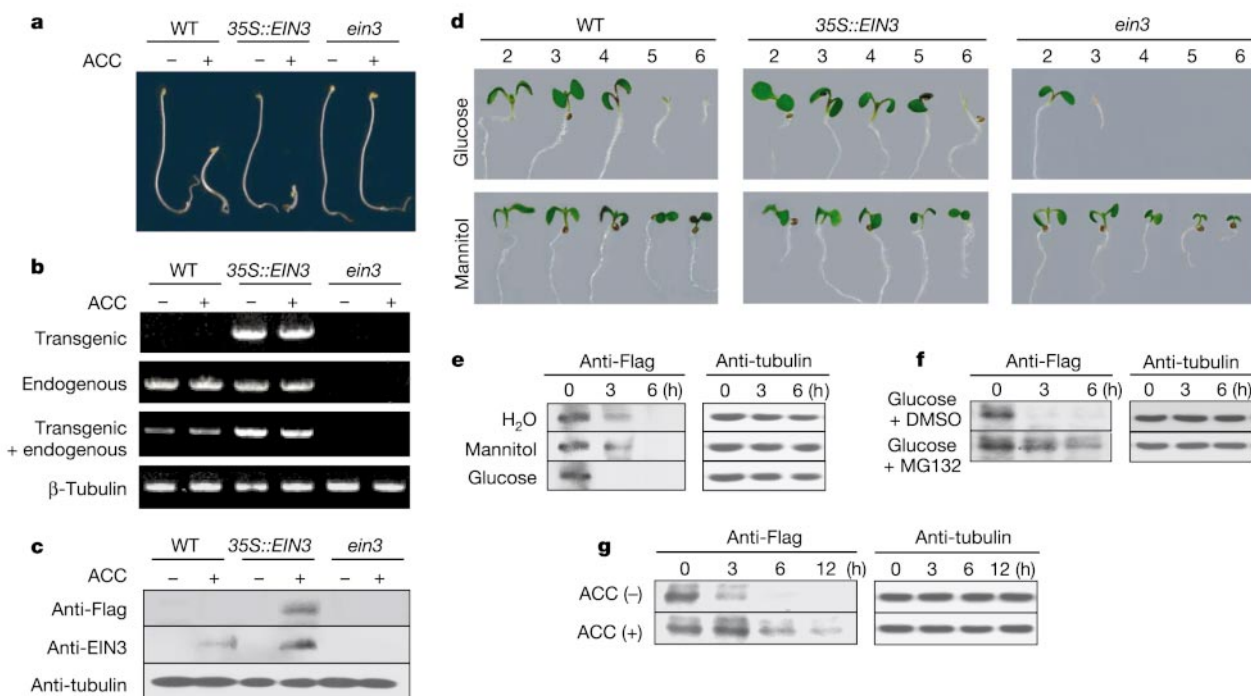


Figure 4 Opposite effects of ethylene and glucose on EIN3 stability in *Arabidopsis*. **a**, Phenotypic analyses of EIN3 transgenic plants. Dark-grown (5 d) wild-type (WT), transgenic (35S::EIN3) and *ein3* seedlings were treated without or with 20 μ M ACC. **b**, Analysis of EIN3 transcripts by PCR with reverse transcription. **c**, Immunoblot analysis of EIN3 protein. **d**, Glucose sensitivity assay. Seedlings were grown under constant light for 6 d. **e**, Glucose-dependent degradation of EIN3. Seedlings were treated with 100 μ M

cycloheximide plus H₂O, 6% mannitol or glucose for the indicated periods. **f**, MG132 inhibits glucose-dependent EIN3 degradation. Seedlings were treated with 100 μ M cycloheximide plus 6% glucose and either dimethylsulphoxide (DMSO) or 20 μ M MG132. **g**, ACC enhances EIN3 stability. Seedlings were treated with 100 μ M cycloheximide plus 20 μ M ACC.

mone signalling, and that mediates auxin²⁵, abscisic acid²⁶, gibberellin²⁷ and brassinosteroid²⁸ responses. Different ubiquitin protein ligases provide some specificity to direct diverse signalling molecules to the proteasome and signalosome^{20,25,29}. It will be interesting to elucidate the precise pathways that link the glucose sensor AtHXK1 to the degradation of EIN3 and mediate the ethylene-enhanced EIN3 stability. Glucose has a profound influence on many essential processes in organisms as diverse as *Escherichia coli*, yeast, mammals and plants, but the complexity of glucose signalling mechanisms in multicellular animals and plants has been recognized only now^{1,3}. Studies of glucose signal transduction in plants may provide insight into glucose signalling mechanisms that are conserved in eukaryotes from yeast to mammals. □

Methods

Effector and reporter constructs

To construct the inducer plasmid, we inserted a chimaeric gene encoding the LexA DNA-binding domain, the nuclear localization signal of the SV40 antigen, the transcriptional activation domain of VP16, and the glucocorticoid receptor (GR) domain¹² between the constitutive 35SC4PPDK promoter (a derivative of the 35S promoter) and the NOS terminator in a general plant expression vector¹³. For dexamethasone-inducible expression of EIN3, the EIN3 cDNA⁶ was fused to a promoter containing four or eight copies of the LexA-binding site. For constitutive expression of EIN3 and EIL1, the respective cDNAs⁶ were fused to sequence encoding the Myc epitope and inserted into the general plant expression vector¹³. The construct expressing constitutively active CTR1 (amino acids 544–799) has been described²³. The inactive CTR1 mutant contained a point mutation (Lys579Met) generated by site-directed mutagenesis. The catalytically inactive AtHXK1 (S177A) has been described⁸. A control reporter construct (NBS–LUC) was generated by fusing the 35S minimal promoter to the firefly LUC gene without any enhancers. We generated the *EBS–LUC* and *GCC–LUC* reporter constructs by inserting four copies each of the EIN3-binding sequence^{7,16} and the GCC-box enhancer sequence²² in front of the 35S minimal promoter.

Protoplast transfection and reporter enzyme assays

Transfection of maize and *Arabidopsis* protoplasts was done as described^{11,15}. As internal controls, we used 35S–GUS or GUS reporter gene fused to the maize or the *Arabidopsis* ubiquitin promoter (*UBI–GUS* and *UBI10–GUS*)^{13,14,16}. The reporter enzyme assays were done as described^{11,15}. Data from duplicate or triplicate samples are shown with error bars. We carried out all reporter enzyme assays at least twice and obtained similar results.

Protoplast *in vivo* labelling and immunoprecipitation analysis

Transfected maize protoplasts were incubated for 1 h for mRNA accumulation before labelling with the [³⁵S]Met/Cys EXPRESS Protein Labeling Mix (NEN) for 5 h. Nuclear and cytoplasmic fractions were separated by treating the labelled protoplasts with 1% triton X-100, followed by centrifugation as described¹⁵. The amount of radioactivity in each sample was verified with a BAS-1000 Bio-imaging Analyzer (Fuji Film). For degradation analysis, transfected protoplasts were labelled for 3 h and then incubated in the presence of a 1,000-fold excess of unlabelled methionine and cysteine. Epitope-tagged proteins were immunoprecipitated with an anti-Myc or anti-HA antibody as described²³, resolved on 12% SDS–PAGE gels, and analysed with the BAS-1000 analyser. For protein immunoblot analyses, we used an anti-Rubisco antibody³⁰, an anti-tubulin antibody (Sigma), a peroxidase-conjugated anti-Flag antibody (Sigma), an anti-HA antibody (Roche), and an anti-EIN3 antibody (a rabbit polyclonal antibody purified with recombinant EIN3 protein expressed from *E. coli*).

GFP fusions and fluorescence microscopy

We produced an expression vector for the GFP fusion protein by replacing the Myc tag with the GFP sequence in the EIN3–Myc vector. The sequence for the N terminus of EIN3 was amplified by polymerase chain reaction (PCR). GFP fluorescence was observed with a TE200 fluorescent microscope (Nikon) or a CK40 microscope (Olympus) equipped with a fluorescence condenser.

Cell-free degradation assay

³⁵S-labelled EIN3–Myc protein was produced in maize protoplasts, purified by immunoprecipitation and then used as a substrate. We prepared cell lysate by grinding maize protoplasts in a buffer supporting proteasome activity²⁰. Degradation reactions were started by mixing the resin trapping the EIN3 or GFP (data not shown) substrate with the cell lysate, and stopped by adding a volume of 2× SDS loading buffer. The samples were separated on a 12% SDS–PAGE gel and analysed with the BAS-1000 analyser.

Transgenic plant analysis

Arabidopsis transgenic plants expressing EIN3–Flag under the control of the 35SC4PPDK promoter were generated by the floral dip method as described^{5,8,15}. All phenotypic and immunoblot analyses were done using T3 homozygous seeds. We obtained similar results with two independent transgenic lines. To detect endogenous and Flag-tagged EIN3, proteins were extracted from 20 seedlings by grinding with 50 µl of 2× SDS loading buffer. For protein degradation analyses, T3 seeds were germinated on MS plates with 20 µM ACC for 5 d in continuous darkness. The seedlings were then incubated in liquid MS medium

supplemented with 100 µM cycloheximide plus the indicated compounds. Proteins extracted from five seedlings by grinding with 20 µl of the 2× SDS-loading buffer were then analysed with an anti-Flag antibody.

Received 26 June; accepted 11 August 2003; doi:10.1038/nature01984.

1. Rolland, F., Winderickx, J. & Thevelein, J. M. Glucose-sensing mechanisms in eukaryotic cells. *Trends Biochem. Sci.* **26**, 310–317 (2001).
2. Smeekens, S. Sugar-induced signal transduction in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 49–81 (2000).
3. Rolland, F., Moore, B. & Sheen, J. Sugar sensing and signaling in plants. *Plant Cell* **14**, S185–S205 (2002).
4. Zhou, L., Jang, J.-C., Jones, T. L. & Sheen, J. Glucose and ethylene signal transduction crosstalk revealed by an *Arabidopsis* glucose-insensitive mutant. *Proc. Natl Acad. Sci. USA* **95**, 10294–10299 (1998).
5. Cheng, W.-H. *et al.* A unique short-chain dehydrogenase/reductase in *Arabidopsis* glucose signaling and abscisic acid biosynthesis and function. *Plant Cell* **14**, 2723–2743 (2002).
6. Chao, Q. *et al.* Activation of the ethylene gas response pathway in *Arabidopsis* by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. *Cell* **89**, 1133–1144 (1997).
7. Solano, R., Stepanova, A., Chao, Q. & Ecker, J. R. Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* **12**, 3703–3714 (1998).
8. Moore, B. *et al.* Role of the *Arabidopsis* glucose sensor HXK1 in nutrient, light and hormonal signaling. *Science* **300**, 332–336 (2003).
9. Chang, C., Kwok, S. F., Blecker, A. B. & Meyerowitz, E. M. *Arabidopsis* ethylene-response gene ETR1: similarity of product to two-component regulators. *Science* **262**, 539–544 (1993).
10. Alonso, J. M. *et al.* EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science* **284**, 2148–2152 (1999).
11. Jang, J.-C. & Sheen, J. Sugar sensing in higher plants. *Plant Cell* **6**, 1665–1679 (1994).
12. Zuo, J., Niu, Q.-W. & Chua, N.-H. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.* **24**, 265–273 (2000).
13. Sheen, J. Protein phosphatase activity is required for light-inducible gene expression in maize. *EMBO J.* **12**, 3497–3505 (1993).
14. Yanagisawa, S. & Sheen, J. Involvement of maize DOF zinc finger proteins in tissue-specific and light-regulated gene expression. *Plant Cell* **10**, 75–89 (1998).
15. Hwang, I. & Sheen, J. Two-component circuitry in *Arabidopsis* cytokinin signal transduction. *Nature* **413**, 383–389 (2001).
16. Kosugi, S. & Ohashi, Y. Cloning and DNA-binding properties of a tobacco ethylene-insensitive3 (EIN3) homolog. *Nucleic Acids Res.* **28**, 960–967 (2000).
17. Chang, C.-C. *et al.* Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. *Plant Cell* **11**, 911–926 (1999).
18. Wang, K. L.-C., Li, H. & Ecker, J. R. Ethylene biosynthesis and signaling networks. *Plant Cell* **14**, S131–S151 (2002).
19. Naidoo, N., Song, W., Hunter-Ensor, M. & Sehgal, A. A role for proteasome in the light response of the timeless clock protein. *Science* **285**, 1737–1741 (1999).
20. Osterlund, M. T. *et al.* Targeted destabilization of HY5 during light-regulated development of *Arabidopsis*. *Nature* **405**, 462–466 (2000).
21. Ravid, T. *et al.* The ubiquitin–proteasome pathway mediates the regulated degradation of mammalian 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *J. Biol. Chem.* **275**, 35840–35847 (2000).
22. Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H. & Ohme-Takagi, M. *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* **12**, 393–404 (2000).
23. Kovtun, Y., Chiu, W.-L., Zeng, W. & Sheen, J. Suppression of auxin signal transduction by a MAPK cascade in higher plants. *Nature* **395**, 716–720 (1998).
24. Kieber, J. J., Rothenberg, M., Gregg, R., Feldmann, K. A. & Ecker, J. R. *CTR1*, a negative regulator of the ethylene response pathway in *Arabidopsis*, encodes a member of the Raf family of protein kinases. *Cell* **72**, 427–441 (1993).
25. Gray, W. M., Kepinski, S., Rouse, D. & Estelle, M. Auxin regulates SCF^{TR1}-dependent degradation of AUX/IAA proteins. *Nature* **414**, 271–276 (2001).
26. Lopez-Molina, L., Mongrand, S. & Chua, N.-H. A postgermination development arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **98**, 4782–4787 (2001).
27. Silverstone, A. L. *et al.* Repressing a repressor: gibberellin-induced rapid reduction of the RGA protein in *Arabidopsis*. *Plant Cell* **13**, 1555–1566 (2001).
28. He, J.-X. *et al.* The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 10185–10190 (2002).
29. Vierstra, R. D. The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci.* **8**, 135–142 (2003).
30. Sheen, J.-Y. & Bogorad, L. Expression of ribulose biphosphate carboxylase large subunit and three small subunit genes in two cell types of maize leaves. *EMBO J.* **13**, 3417–3422 (1986).

Acknowledgements We thank B. Lam for help with antibody preparation and for providing plants; T. Miwa for help with generating transgenic *Arabidopsis*; M. V. Parthasarathy for the actin antibody; F. Rolland and B. Seed for critically reading the manuscript; J. Callis for the *UBQ10–GUS* construct; B. Moore for the AtHXK1 mutant; Y. Niwa for the *TP–GFP* construct; and Y.-H. Cho and R. Patharkar for advice. The work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to S.Y.), and the National Science Foundation and the National Institutes of Health (to J.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.Y. (yanagi-s@rib.okayama-u.ac.jp).

Role of visual pigment properties in rod and cone phototransduction

Vladimir Kefalov^{1*}, Yingbin Fu^{2*}, Nicholas Marsh-Armstrong^{1,3} & King-Wai Yau^{1,2}

¹Department of Neuroscience, ²Howard Hughes Medical Institute, and ³Kennedy-Krieger Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

* These authors contributed equally to this work

Retinal rods and cones share a phototransduction pathway involving cyclic GMP¹. Cones are typically 100 times less photosensitive than rods and their response kinetics are several times faster², but the underlying mechanisms remain largely unknown. Almost all proteins involved in phototransduction have distinct rod and cone variants. Differences in properties between rod and cone pigments have been described, such as a 10-fold shorter lifetime of the meta-II state (active conformation) of cone pigment^{3–6} and its higher rate of spontaneous isomerization^{7,8}, but their contributions to the functional differences between rods and cones remain speculative. We have addressed this question by expressing human or salamander red cone pigment in *Xenopus* rods, and human rod pigment in *Xenopus* cones. Here we show that rod and cone pigments when present in the same cell produce light responses with identical amplification and kinetics, thereby ruling out any difference in their signalling properties. However, red cone pigment isomerizes spontaneously 10,000 times more frequently than rod pigment. This high spontaneous activity adapts the native cones even in darkness, making them less sensitive and kinetically faster than rods. Nevertheless, additional factors are probably involved in these differences.

Human or salamander red cone pigment, together with green fluorescent protein (GFP) for facilitating screening, was introduced as a transgene into *Xenopus* under the control of the cytomegalovirus (CMV) promoter. In a wild-type or GFP control *Xenopus* retina, an antibody to red cone pigment labelled only the outer segments of sporadic red cones. In a frog expressing transgenic red cone pigment, however, immunolabelling included the abundant rod outer segments (Fig. 1a). The concentrated labelling of rod outer segments suggested that the localization signal that targets cone pigment to the outer segment is also recognized by rods. The immunostaining was usually non-uniform across the whole retina, suggesting variable expression of the transgenic pigment among the rods (see below).

Outer-segment membrane current was recorded from single principal ('red') rods with a suction pipette. Rods expressing transgenic cone pigment showed no marked changes in their flash responses (Fig. 1b), but their sensitivity was, on average, half that of wild-type or GFP control rods (Fig. 2a and Table 1). In darkness, these rods also showed considerably higher current noise, which was suppressible by light (Fig. 1c). The photosensitivity of the noise, together with its kinetic characteristics (see below), suggested that it originated in the phototransduction pathway. In other control experiments with transgenic human rod pigment expressed in *Xenopus* rods, no increase in dark noise was observed (data not shown).

The increase in dark noise in rods expressing transgenic red cone pigment suggested that cone pigment was more prone to spontaneous isomerization than rod pigment, and that it was coupled functionally to the rod phototransduction pathway. Such coupling was confirmed by a small red shift in the spectral sensitivity of rods expressing cone pigment (Fig. 2a, c), caused by a $\lambda_{\max}$ of ~ 620 nm for red cone pigment as compared with ~ 520 nm for rod pigment^{9–11}, with 11-*cis*-dehydroretinal (vitamin A2) as the chro-

mophore (which *Xenopus* retina has almost exclusively; Fig. 2a, inset, and Methods).

On the basis of the average expression of cone pigment described below, 520-nm light should stimulate essentially only rod pigment, whereas 700-nm light should stimulate cone pigment with $\sim 25\%$ probability (31% for the cell in Fig. 2b). In a series of identical dim flashes, the averaged response waveform were unexpectedly identical at both wavelengths (Fig. 2b, d, left), indicating that the responses triggered by rod and cone pigments had the same kinetics.

The response amplitude triggered by a single pigment molecule, derived from the response ensemble variance-to-mean ratio in the dim-flash experiment, was also very similar at 520 and 700 nm

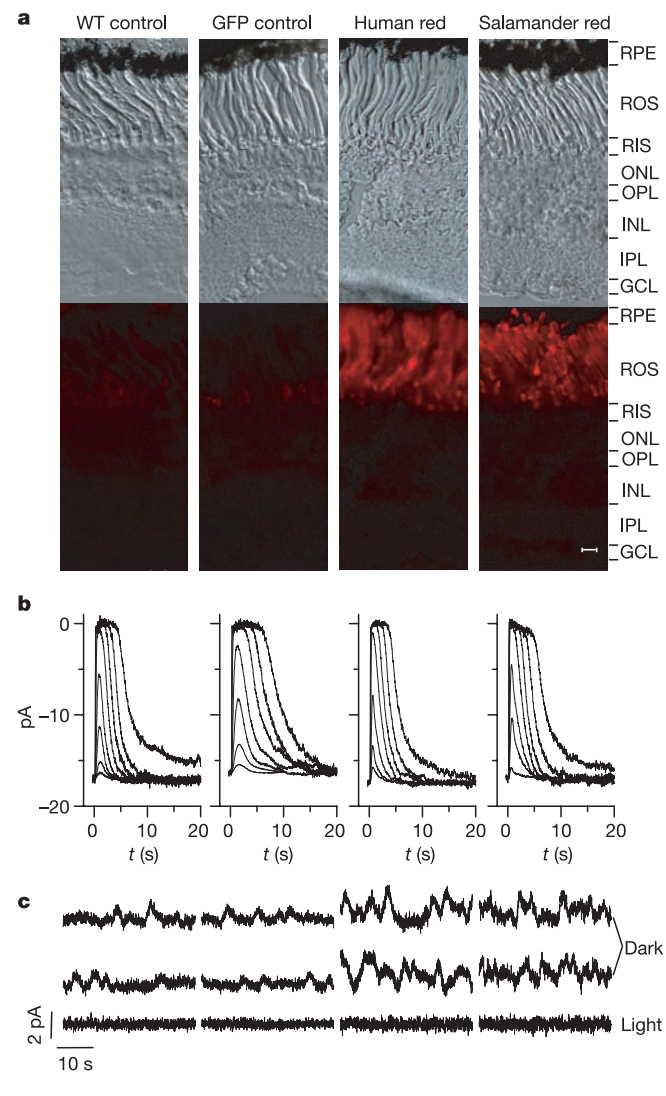


Figure 1 *Xenopus* rods expressing transgenic human and salamander red cone pigments. **a**, Frozen sections of *Xenopus* retinas immunostained for red cone pigment. Transgenic frogs expressed GFP alone (GFP control), or together with human or salamander red cone pigment. RPE, retinal pigment epithelium; ROS, rod outer segment; RIS, rod inner segment; ONL, outer nuclear layer; OPL, outer plexiform layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglion cell layer. Scale bar, 10 μ m. **b**, Flash intensity response families from rods of the genotypes in **a**. In these suction-pipette recordings, a 20-ms flash at time 0 delivered 0.12, 0.41, 1.55, 5.19, 25.6, 85.9, 328 or 1,100 (520 nm) photon μ m⁻². The lowest flash intensity shown is 0.12 photon μ m⁻² for wild-type and GFP rods and 0.41 photon μ m⁻² for transgenic rods expressing human or salamander red cone pigment. **c**, Dark-current noise in control and transgenic rods. Note the increased noise in rods expressing cone pigment ('dark' traces in the two right panels). The 'light' trace was recorded in saturating light, which delivered 1.52×10^7 photon μ m⁻² s⁻¹ (520 nm).

(~ 0.61 pA versus 0.56 pA; Fig. 2b and Table 1). The variance-to-mean ratio at 700 nm does not have a simple meaning because the response ensemble consists of mixed responses from rod and cone pigments. However, computer simulations of two mixed-response populations, having various amplitude ratios and occurring according to the Poisson distribution with realistic parameters to mimic our experiments (Supplementary Information), suggested that the responses produced by single rod and cone pigment molecules were unlikely to differ by more than twofold in amplitude. A more definitive analysis based on the dark-current noise indicated that the two amplitudes are essentially identical (see below).

The percentage of transgenic cone pigment in host rods was actually very small. Given the identical responses produced by rod and cone pigments, this percentage can be estimated by fitting a composite absorption spectrum to the measured action spectrum, especially at long wavelengths. This gave $0.030 \pm 0.005\%$ ($n = 17$) for human and $0.033 \pm 0.017\%$ ($n = 12$) for salamander red cone pigment (Fig. 2a, c), which is broadly comparable to the expression of transgenic rod pigment in *Xenopus* rods¹². These estimates are insensitive to uncertainties of ± 3 nm in the $\lambda_{\max}$ values for the spectral templates of the two pigments.

We also generated *Xenopus* expressing a mutant human red cone

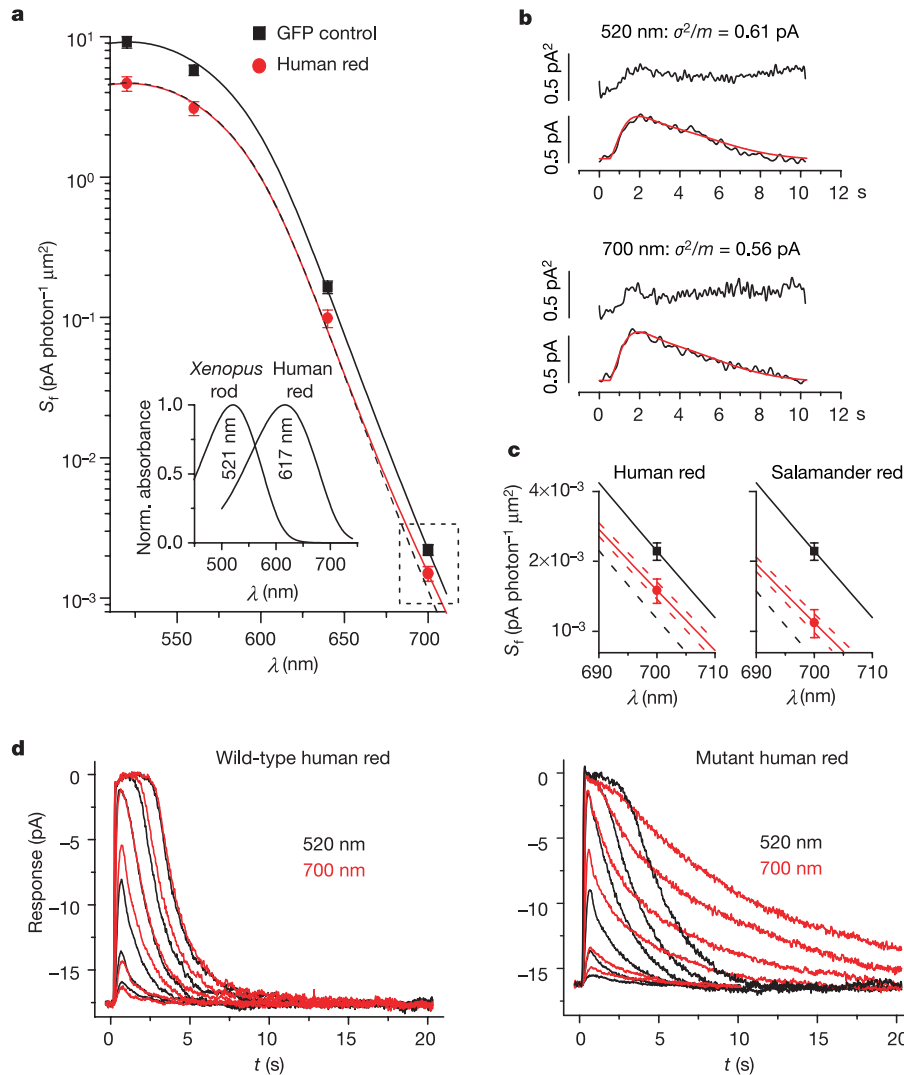


Figure 2 Spectral shift and dim-flash responses in rods expressing transgenic cone pigment. **a**, Red shift induced by human red cone pigment. Averaged flash sensitivity (S_f) for rods expressing GFP alone ($n = 15$) was twice that of those expressing GFP plus human red cone pigment ($n = 17$). S_f was derived from a series of identical dim flashes at 520, 560, 640 and 700 nm. Error bars indicate the s.e.m. Data are fit by the A2 spectral template for *Xenopus* 'rod' pigment (black) and for a combination of 99.970% *Xenopus* rod pigment and 0.030% human red cone pigment A2 spectral templates (red). The broken black trace is identical to the unbroken one but shifted vertically (to compensate for the desensitization of rods expressing cone pigment) for comparison with the red trace. Note the increased sensitivity at long wavelengths of rods expressing cone pigment. Inset, A2 spectral templates for *Xenopus* 'rod' pigment ($\lambda_{\max} = 521$ nm) and human red cone pigment ($\lambda_{\max} = 617$ nm; Methods). **b**, Comparison of dim-flash responses at 520 and 700 nm for a rod expressing human red cone pigment. Repeated dim flashes delivered 4.1×10^{-1} and 1.34×10^3 photon μm^{-2} at 520 and 700 nm, respectively. The effective collecting area of this cell was unusually small ($4.2 \mu\text{m}^2$).

According to the spectral templates, the two wavelengths should discriminate for one or the other pigment by 2.68-fold and 1,133-fold, respectively. With the percentage of transgenic cone pigment in this cell estimated at 0.039%, 31% of the flash responses at 700 nm should have been triggered by cone pigment. The same mathematical waveform (red traces) fits the mean responses at both wavelengths, indicating identical kinetics. The ratio between σ^2 (top black traces) and m (bottom black traces) at response peak is indicated. In computing σ^2 , the beginning of each response trace (before the flash) was artificially superposed; because of the high dark noise, this caused the progressive drift (increase) in the variance trace. **c**, Left, magnification of boxed region in **a** showing resolution of the pigment ratio. Broken red lines are drawn at $0.030 \pm 0.010\%$ cone pigment. Right, similar averaged data for salamander red cone pigment. Unbroken red line is drawn at 0.033%, and broken red lines at $0.033 \pm 0.010\%$, cone pigment. **d**, Effect of mutating the putative phosphorylation sites of human red cone pigment on rod response kinetics. The intensity of flashes delivered at time 0 are given in the Methods.

Table 1 Comparison of parameters for wild-type *Xenopus* rods and *Xenopus* rods expressing transgenic cone pigment

	WT control (<i>n</i> = 9)	GFP control (<i>n</i> = 15)	Human rod pigment (<i>n</i> = 9)	Human red cone pigment (<i>n</i> = 17)	Salamander red cone pigment (<i>n</i> = 12)
Dark current (pA)	10.5 ± 0.5	16.8 ± 1.2	15.0 ± 0.5	13.2 ± 1.2	11.9 ± 0.9
Single-photon response (pA)	0.55 ± 0.03 (520 nm)	0.61 ± 0.05 (520 nm)	0.55 ± 0.03 (520 nm)	0.48 ± 0.02 (520 nm) 0.48 ± 0.03 (700 nm)	0.35 ± 0.04 (520 nm) 0.30 ± 0.06 (700 nm)
Time to peak (s)	1.24 ± 0.04 (520 nm)	1.32 ± 0.04 (520 nm)	1.22 ± 0.05 (520 nm)	1.03 ± 0.04 (520 nm) 1.03 ± 0.04 (700 nm)	1.02 ± 0.06 (520 nm) 0.90 ± 0.07 (700 nm)
Flash sensitivity at 520 nm (photons ⁻¹ μm ²) [*]	0.46 ± 0.02	0.55 ± 0.04	0.54 ± 0.05	0.36 ± 0.02	0.30 ± 0.04
Thermal isomerization rate (<i>R</i> [*] s ⁻¹) [†]	0.17 ± 0.03	0.17 ± 0.02	0.06 ± 0.01	0.73 ± 0.16	0.77 ± 0.27

^{*}Flash sensitivity values have been normalized with respect to the dark current. All values are the mean ± s.e.m.

[†]Raw rates derived from power spectral fits. For control rods, the continuous component of the dark noise has not been subtracted. For rods expressing transgenic pigment, the noise observed in control rods has not been subtracted (see text).

pigment lacking all putative phosphorylation sites in the carboxy terminus to remove any inactivation by pigment kinase, and found that the host rods showed longer-lasting responses at 700 nm than at 520 nm (Fig. 2d, right). This experiment verifies that transgenic red cone pigment is truly functional and that, despite the rapid decay of the meta-II state of red cone pigment, its shutoff is still dominated by phosphorylation and arrestin binding preceding the decay, as it is for rod pigment (see also ref. 13). Because the responses produced by endogenous rod pigment and transgenic wild-type red cone pigment have identical waveforms, the pigment kinase and arrestin in rods must act identically on rod and cone pigments (Supplementary Information).

We analysed the dark-current noise in rods by computing its power density spectrum. For GFP control and wild-type rods, this power spectrum (after subtracting the power spectrum of residual noise in saturating light) agreed fairly well with that derived from the measured single-photon response of endogenous rod pigment (Fig. 3a), as expected¹⁴. The fit gave an average spontaneous isomerization (*R*^{*}) rate for the rod pigment of $0.17 \pm 0.02 R^* s^{-1}$ (*n* = 15) in GFP control rods ($0.14 R^* s^{-1}$ for the cell in Fig. 3a), and $0.17 \pm 0.03 R^* s^{-1}$ (*n* = 9) in wild-type rods. A correction has to be made to remove a 'continuous' component¹⁴ of the dark noise originating from intrinsic phosphodiesterase activity¹⁵. Assuming this continuous noise to be comparable in variance to the discrete

noise¹⁴, the power density for the 'discrete' noise, reflecting pigment isomerization only, should be roughly halved, corresponding to $\sim 0.1 R^* s^{-1}$ (at 21–23 °C).

For rods expressing transgenic red cone pigment, the power spectrum of the much higher dark noise was also fit well by that derived from the single-photon response of rod pigment triggered with dim 520-nm flashes (Fig. 3b). This fit confirmed that the response kinetics were indeed identical whether generated by rod or cone pigment. Given that the responses produced by single rod and cone pigment molecules had essentially identical amplitudes (see below), the fit gave an average spontaneous isomerization rate of $0.73 \pm 0.16 R^* s^{-1}$ (*n* = 17) for cells expressing human red cone pigment ($1.4 R^* s^{-1}$ for the cell in Fig. 3b) and $0.77 \pm 0.27 R^* s^{-1}$ (*n* = 12) for cells expressing salamander red cone pigment. Subtracting all endogenous dark noise in the control rods described above gave rates for human and salamander cone pigments of $0.56 R^* s^{-1}$ and $0.60 R^* s^{-1}$, respectively. The net rate in individual cells ranged from 0 to $2.8 R^* s^{-1}$, a variation attributed to the variable expression of cone pigment, as suggested by immunostaining (see above). Indeed, the calculated net rate in each cell was roughly proportional to the estimated percentage of cone pigment (Fig. 3c).

The low expression of the transgenic cone pigment actually has an advantage. In three rods expressing cone pigment, the increase in dark noise, although substantial, showed occasional breaks to reveal

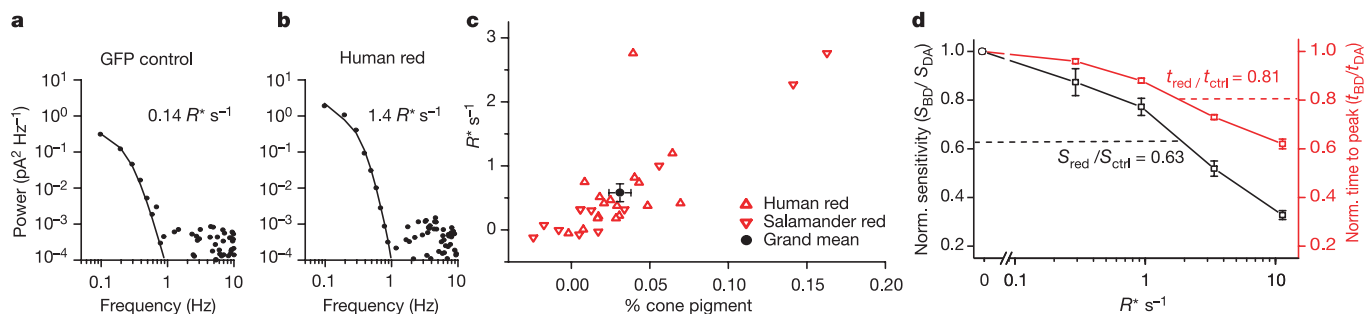


Figure 3 Analysis of dark-current noise. **a**, Power density spectrum of the dark noise from a GFP control rod (filled circles) fit by the power spectrum computed from the single-photon response from endogenous rod pigment (continuous trace). The derived rate of thermal isomerization for this cell was $0.14 R^* s^{-1}$. **b**, As **a**, but for a rod expressing human red cone pigment. The derived rate of thermal isomerization was $1.4 R^* s^{-1}$. **c**, Collected plot of the net rate of thermal isomerization of transgenic cone pigment ($R^* s^{-1}$) against the estimated percentage of cone pigment in individual host rods (open red symbols). The average dark noise of control rods ($0.17 R^* s^{-1}$) has been subtracted. Cone pigment expression was calculated as in Fig. 2. The negative percentages for several cells with essentially undetectable expression were due to measurement scatter, which made the measured flash sensitivity at 700 nm fall below the absorption spectrum for the endogenous rod pigment. Black circle shows grand mean ± s.e.m. for the

transgenic rods expressing human or salamander red cone pigments. **d**, The decrease in flash sensitivity and reduction in 'time to peak' of the dim-flash response due to transgenic red cone pigment could be reproduced in GFP control rods by a background light producing $1\text{--}2 R^* s^{-1}$. Abscissa, intensity of steady background light on GFP control rods. Left ordinate, ratio of dim-flash sensitivities in the presence (S_{BD}) or absence (S_{DA}) of background light for GFP control rods. Right ordinate, ratio of time to peak of dim-flash responses in the presence (t_{BD}) or absence (t_{DA}) of background light for GFP control rods. Calculation of the ratios S_{BD}/S_{DA} and t_{BD}/t_{DA} is described in Methods. Note that both ratios intersect the background-light data at almost the same background light intensity, as expected. This value of $\sim 1.8 R^* s^{-1}$ is broadly similar to the spontaneous isomerization rate estimated from power spectral analysis of the dark noise in rods expressing transgenic cone pigment.

the quiet dark-current baseline (for example, Fig. 1c, third column, top trace). In these cases, an objective estimate of the unitary amplitude (that is, the response amplitude triggered by a single isomerized cone pigment molecule) underlying the noise can be calculated from $\sigma^2/m \times [\int f(t)dt / \int f^2(t)dt]$ (ref. 16), where σ^2 is the steady-current variance, m is the mean current relative to this baseline, and $f(t)$ is a function normalized to unity at peak and with a power spectrum that fits the dark-noise spectrum.

Before calculation, the σ^2 and m components reflecting spontaneous isomerization of endogenous rod pigment (σ_r^2 and m_r) were evaluated and subtracted from the overall σ^2 and m values, using $\sigma_r^2 = \nu_r \alpha_r^2 \int f^2(t)dt$ and $m_r = \nu_r \alpha_r \int f(t)dt$ (ref. 16), where ν_r is the average isomerization frequency ($0.1 R^* s^{-1}$, as measured above) and α_r is the response peak amplitude triggered by a single rod pigment molecule measured with dim 520-nm flashes (Supplementary Information).

The response triggered by a single cone pigment molecule derived in this way, 0.55 ± 0.01 pA ($n = 3$), agreed well with the rod pigment single photon response of 0.49 ± 0.02 pA measured in the same cells with dim 520-nm flashes. This agreement indicates that rod and cone pigments indeed generate responses of essentially identical size. With the unitary amplitude known, the rate of cone pigment events could be calculated from either of the above equations and was determined to be $1.13 \pm 0.16 R^* s^{-1}$, which becomes $1.23 \pm 0.16 R^* s^{-1}$ after adding ν_r . This value is close to

the value of $1.09 \pm 0.14 R^* s^{-1}$ derived by power spectral analysis in the same three cells.

The high thermal isomerization rate of the transgenic cone pigment was equivalent to a steady background light and should adapt the host rod even in darkness. This adaptation could explain why rods expressing cone pigment were less sensitive than normal. Indeed, the average sensitivity ratio between transgenic and control rods in darkness (0.63; Fig. 3d) could be reproduced in control rods by exposing them to a steady light of $1.8 R^* s^{-1}$, which is broadly similar to the equivalent light derived from the noise analysis described above. The same conclusion can be drawn by using the response 'time to peak' as the parameter for comparing transgenic and control rods (Fig. 3d).

Separately, by using the cones of *Xenopus* expressing transgenic human rod pigment, we could study the converse situation, that is, rod pigment functioning in a cone environment. We chose the abundant 'red' cones for recording. The functional coupling of transgenic rod pigment to the host cone transduction pathway was detected by a blue shift in the action spectrum (Fig. 4a). The fit of the measured action spectrum by a linear combination of the A2 absorption templates for human rod pigment and *Xenopus* cone pigment (Fig. 4b and Methods) indicated that, on average, ~12% of the action spectrum was contributed by transgenic rod pigment. Why a transgenic pigment is expressed at a considerably higher

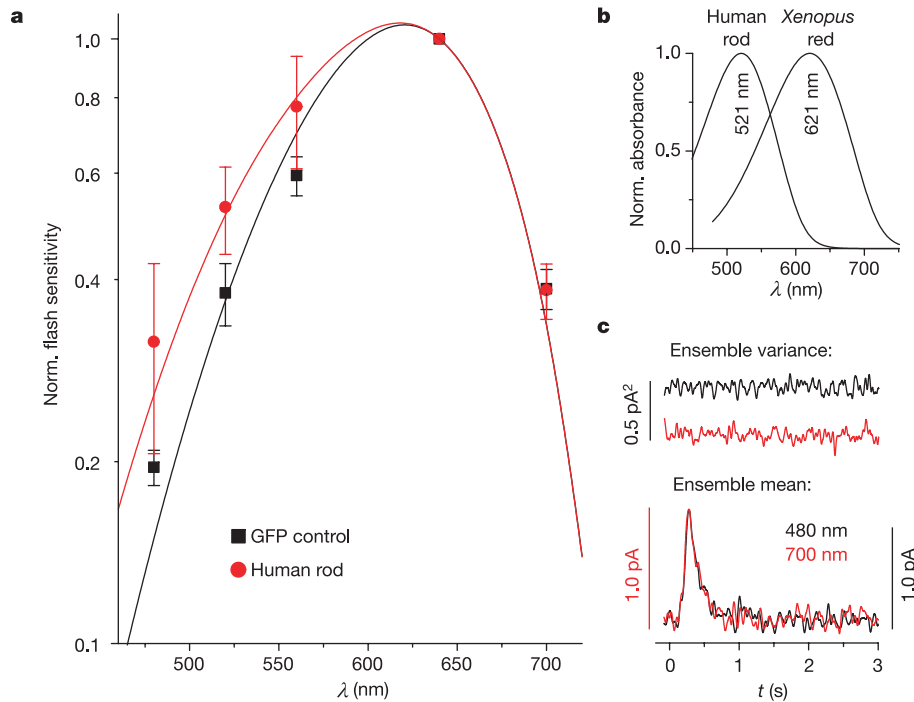


Figure 4 *Xenopus* red cones expressing transgenic human rod pigment. **a**, Blue shift in action spectrum. S_f from cones expressing GFP alone or together with human rod pigment was averaged ($n = 10$ each) and normalized at 640 nm. Error bars indicate the s.e.m. The data are fit by the A2 spectral template for *Xenopus* red cone pigment (black) and for a combination of 88% contribution from the A2 template of *Xenopus* red cone pigment and 12% contribution from the A2 template of transgenic human rod pigment (red). Note the increased sensitivity at short wavelengths for cones expressing rod pigment. In principle, S_f at 560 nm ought to be the same in the absence and presence of transgenic rod pigment because this is near the isosbetic point (see **b**). There is probably a slight uncertainty in the measurements at these wavelengths, because the two sets of data matched well at long wavelengths (consistent with the dominance of the contribution from the endogenous cone pigment in this spectral region). Constraining a sensitivity invariance at 560 nm would give a contribution of 2.7% (instead of 12%) from the A2 template of the human rod pigment. **b**, A2 spectral templates for *Xenopus* red cone pigment ($\lambda_{max} = 621$ nm) and human rod pigment ($\lambda_{max} = 521$ nm; Methods). At

480 nm, the rod pigment would be preferentially stimulated 5.4-fold; at 700 nm, the cone pigment would be preferentially stimulated 1,322-fold. **c**, Identical kinetics of dim-flash responses at 480 and 700 nm for a *Xenopus* cone expressing transgenic human rod pigment. The scales for the two responses have been adjusted so that the mean response at 480 nm (actual amplitude, 1.11 pA) matches that at 700 nm (0.93 pA). Responses represent averages from 30 flash trials each. At each wavelength, the ensemble variance of the response was negligible, indicating that the unitary response was extremely small. The percentage of transgenic rod pigment was estimated to be 6.6% in this cell. Because the probability of absorption at 480 nm is 5.4-fold higher for human rod pigment than for *Xenopus* red cone pigment, it can be calculated that ~28% of the response at 480 nm should have been triggered by rod pigment. The single-photon response triggered by the endogenous red cone pigment is about 0.008 pA (Methods), consistent with the extremely small variance of the flash responses. This value is about 70 times less than that for the single-photon response of a *Xenopus* red rod.

percentage in cones than in rods is not clear at present. Again, dim-flash responses elicited at 480 and 700 nm showed identical kinetics (Fig. 4c, bottom).

From the above ~12% and the absorption templates, the transgenic rod pigment should contribute, on average, ~42% of the light response at 480 nm (28% for the cell in Fig. 4c) but only 0.01% at 700 nm. Thus, the kinetic invariance at the two wavelengths strongly suggested that transgenic rod pigment and endogenous cone pigment produced identical response waveforms. At the same time, there was no detectable increase in ensemble variance during the dim-flash response at 480 or 700 nm (Fig. 4c, top), consistent with extremely small unitary amplitudes in both cases and therefore ruling out the possibility that a much larger response was produced by transgenic rod pigment than by endogenous cone pigment. These conclusions mirrored those derived above from rods expressing transgenic cone pigment.

In summary, rod and cone pigments are completely equivalent with respect to signalling downstream in phototransduction: first, the active lifetimes of both are dictated by shutoff regulated by phosphorylation and arrestin binding rather than by meta-II decay; second, the pigment kinase and arrestin in a rod or cone act identically on rod and cone pigments; and third, rod and cone pigments have identical efficacies when coupled to a given (rod or cone) transducin. Nonetheless, the two do differ in that cone pigment (or at least red cone pigment) is intrinsically much more prone to spontaneous isomerization.

How large is this difference? The spontaneous isomerization rate of rod pigment in control *Xenopus* rods is $\sim 0.1 R^* s^{-1}$ (see above), corresponding to a molecular rate constant of $\sim 3.7 \times 10^{-11} s^{-1}$ (Methods). Similar calculations give a molecular rate constant of $\sim 6.7 \times 10^{-7} s^{-1}$ for salamander red cone pigment (1.8×10^4 -fold higher). Applying this rate constant to the native salamander red cone (Methods) gives $\sim 175 R^* s^{-1}$. This rate is 3.5–7.0 times lower than that previously estimated from native salamander red cones^{8,17}. For human red cone pigment, the disparity is even larger between our estimated rate and that measured from native monkey red cones (Supplementary Information). These disparities suggest that a significant portion of the dark noise in a native cone does not originate from the pigment.

It is possible that, in native cones, spontaneous transducin activity contributes to the noise. In addition, although some spontaneous transduction activity in cones is dependent on GTP⁸ and therefore triggered by either pigment or transducin, not all dark transduction activity has to originate exclusively from these two sources, but instead can come from spontaneous phosphodiesterase activity (Supplementary Information). Finally, some dark noise in native cones previously ascribed to the transduction process^{8,17} might not originate from the outer segment at all. Unlike the rod cGMP-gated channel, which behaves as a constant-current generator by showing a flat current–voltage relation at physiological membrane potentials¹, the cone cGMP-gated channel shows substantial slope conductance at these potentials¹ and therefore can be influenced by voltage noise at the inner segment or cell body. This noise is also light sensitive, because the cGMP-gated channels are closed in the light.

How much adaptation can be brought about by the spontaneous isomerization activity of cone pigment? For salamander red cones, the equivalent light of $175 R^* s^{-1}$ will desensitize these cells six- to sevenfold, which accounts for the bulk of the overall sensitivity difference between salamander rods and red cones (14- to 15-fold^{8,18}, note that folds multiply; Supplementary Information). In this regard, it is interesting to note that salamander green rods and blue cones share the same pigment and also have remarkably similar single-photon response amplitudes and kinetics¹⁹. Although this sharing of the same pigment by classes of rods and cones is rather unusual (for example, it is not found in primates), the situation nonetheless suggests that the phototransduction components down-

stream from pigment may not have much of a role in determining the difference in absolute sensitivity between rods and cones.

In primates², rods and red cones differ in sensitivity by ~100-fold, perhaps half of which (~10-fold) is possibly accounted for by the spontaneous isomerization of red cone pigment (Supplementary Information). The same conclusion can be drawn from published data for fish²⁰, assuming a comparable rate constant for fish and salamander red cone pigment. Thus, in general, the high rate of spontaneous isomerization of the red cone pigment seems to be an important factor underlying the difference in absolute sensitivity between rods and red cones in darkness. At the same time, additional mechanisms are probably involved. These other factors could be quantitative differences in the functional coupling between transduction proteins acting downstream from the pigment, which might produce different amplifications at the respective stages²¹ and also quantitatively different Ca^{2+} feedbacks^{18,22}. Biochemical experiments in fish have indicated that such differences do exist²¹, but so far these experiments have been done at very low concentrations of free Ca^{2+} (calculated as ~30 nM or lower) and therefore might reflect merely a highly adapted state, which may well differ between rods and cones. The issue on downstream components remains to be elucidated. Finally, this study deals only with the red pigment, and it will be important to examine the green and blue pigments as well. □

Methods

Generation of transgenic frogs

All constructs were prepared in pCS2+-based vectors, designed for optimal protein expression in *Xenopus*. The expression of pigment and GFP genes was controlled by a CMV promoter, which can drive transgene expression in both rods and cones of *Xenopus*. In generating the (presumably) phosphorylation-defective mutant of human red cone pigment, all ten putative phosphorylation sites (S345, S348, S349, S351, T353, S356, S357, S359, S360 and S362) in the C terminus were changed to alanine. All cloning and mutagenesis were done with standard methods. We produced transgenic *Xenopus* embryos by using restriction-enzyme-mediated integration as described²³. Hatched frogs were kept under a 12 h/12 h light/dark cycle before use.

Immunocytochemistry

Fresh eyecups from a pithed frog were fixed overnight at 4 °C in PBS containing 4% paraformaldehyde. After further overnight incubation in PBS plus 30% sucrose at 4 °C, the eyecups were embedded in OCT (Tissue-Tek) and sectioned at 10–20 µm. The retinal sections were processed for immunocytochemistry with polyclonal antibody JN492 (ref. 24) specific for red cone pigment, followed by Cy3-conjugated goat anti-rabbit IgG. Despite the use of the promiscuous CMV promoter, the immunolabelling was predominantly located in the rod outer segments, which could indicate particularly strong transgene expression in rods combined with specific targeting, or perhaps an instability of the transgenic pigment protein in other retinal cell types. By contrast, GFP expressed under the CMV promoter was much more widespread in the retina (data not shown).

Suction pipette recordings

Small frogs (beyond stage 66, the peak of metamorphosis; 2–5 cm long) dark-adapted overnight were double-pithed and the eyes were removed and hemisected. The retina was peeled off and chopped, and the pieces were placed in a recording chamber and perfused with Ringer solution containing 110 mM NaCl, 2.5 mM KCl, 1.6 mM MgCl₂, 1.0 mM CaCl₂, 10 mM dextrose, 10 mM HEPES and 0.1 µg ml⁻¹ bovine serum albumin (pH 7.8). Under infrared illumination, the outer segment of a rod or cone photoreceptor projecting from a retinal fragment was drawn into a tight-fitting glass pipette for recording and stimulated with calibrated 20-ms flashes or bright steady light. The signals were low-pass filtered at 20 Hz (8-pole Bessel) and digitized at 100 Hz for further analysis. For power spectral analysis, the signals were low-pass filtered at 20 Hz (8-pole Butterworth) and digitized at 100 Hz.

To determine the effect of mutating the phosphorylation sites in human red cone pigment (Fig. 2d), flashes delivered 0.41, 1.55, 5.19, 25.6, 85.9 and 328 photon µm⁻² at 520 nm, and 1.38×10^3 , 4.38×10^3 , 2.33×10^4 , 7.41×10^4 , 2.87×10^5 and 9.13×10^5 photon µm⁻² at 700 nm for rods expressing the wild-type cone pigment; and 0.35, 1.35, 4.54, 22.4, 75.2 and 287 photon µm⁻² at 520 nm, and 1.52×10^3 , 4.82×10^3 , 2.56×10^4 , 8.15×10^4 , 3.15×10^5 and 1.0×10^6 photon µm⁻² at 700 nm for rods expressing the mutated cone pigment.

Single-photon response and power spectral analysis

The amplitude of the single-photon response was calculated as the ensemble amplitude variance-to-mean ratio of the transient peak of the responses to a series of 30–100 identical dim flashes. In experiments involving power density spectra, the average response to a dim-flash series was scaled to the single-photon response amplitude and then fit by a function consisting of the convolution of four single exponentials with optimally adjusted time constants. This function was used for computing the power spectrum of the single-photon response. Power spectra of the noise in both darkness and bright light were calculated using Clampfit8 (Axon Instruments) in 10.24-s segments with 50% overlap,

from ten sweeps of 40.96 s each. The dark-minus-light difference spectrum was obtained by subtraction. The rate of thermal isomerization of the visual pigment was estimated by scaling the power spectrum of the single-photon response to fit the difference spectrum and dividing this scaling factor by the acquisition time (10.24 s).

For the response kinetics of *Xenopus* cones expressing human rod pigment (Fig. 4c), the single-photon response was calculated as follows. The flash intensity at 700 nm delivered the equivalent of 1.37×10^3 photon μm^{-2} at 621 nm (λ_{max} for endogenous red cone pigment). With an average effective collecting area of $0.08 \mu\text{m}^2$ for the *Xenopus* red cone outer segment (length $4.3 \mu\text{m}$ by mean diameter $1.6 \mu\text{m}$, unpolarized illumination incident transversely at the outer segment, assuming a transverse optical density of $0.012 \mu\text{m}^{-1}$ and a quantum efficiency of isomerization of 0.67), this intensity should produce about 110 cone pigment isomerizations. The single-photon response triggered by the endogenous red cone pigment was therefore about 0.008 pA (mean response at 700 nm (0.93 pA) divided by 110).

Pigment absorption spectrum templates

We adopted published visual pigment templates¹¹ for describing the absorption spectra of the various pigments. The best fit of the A2 template to our spectral sensitivity measurements from GFP control rods (Fig. 2a) gave $\lambda_{\text{max}} = 521$ nm for the *Xenopus* red rod pigment, consistent with previous reports^{9–11}. For salamander red cone pigment with A2 chromophore, we used $\lambda_{\text{max}} = 620$ nm (ref. 25). For the polymorphic form of the human red cone pigment used here, $\lambda_{\text{max}} = 557$ nm with A1 chromophore²⁶ was adopted. We used an empirical expression²⁷ to derive its corresponding λ_{max} with A2 chromophore and obtained $\lambda_{\text{max}} = 617$ nm. For the pigment of the principal *Xenopus* 'red' cones, the fit of our data to the A2 template¹¹ gave $\lambda_{\text{max}} = 621$ nm, which is slightly higher than reported⁹. For human rod pigment, $\lambda_{\text{max}} = 498$ nm (ref. 28) with A1 chromophore would give $\lambda_{\text{max}} = 521$ nm with A2 chromophore²⁹.

The small amount (a few per cent) of A1 pigment presumably present in the *Xenopus* retina¹⁰ was ignored in this work. Its inclusion would produce a negligible difference in the estimated expression of the transgenic pigment.

Pigment content of cells

In our experiments, the *Xenopus* rod outer segment had a mean diameter of $6.4 \mu\text{m}$ and, on average, $40 \mu\text{m}$ of the outer segment length was recorded by the suction pipette. Thus, the recorded outer segment volume was $1.3 \times 10^3 \mu\text{m}^3$. Adopting a standard pigment concentration of 3.5 mM (ref. 30) in an outer segment, this volume corresponds to 2.7×10^9 pigment molecules. The corresponding number of cone pigment molecules in a transgenic *Xenopus* rod can be calculated by simply multiplying the rod pigment number with the corresponding percentage of total pigment content derived from the red-shift experiments (0.030% for human red cone pigment and 0.033% for salamander red cone pigment). For salamander red cones, its outer segment is, on average, $4 \mu\text{m}$ in diameter by $10 \mu\text{m}$ in length¹⁹. At 3.5 mM (ref. 30), the total cone pigment content is therefore 2.6×10^8 molecules. These numbers are used for calculating pigment thermal rate constants or the overall rate of thermal isomerization in a cell.

Calculation of flash sensitivity and time-to-peak ratios

In Fig. 3d, the grand-mean dim-flash sensitivity ratio ($S_{\text{red}}/S_{\text{ctrl}}$) and response 'time to peak' ratio ($t_{\text{red}}/t_{\text{ctrl}}$) of rods with or without transgenic cone pigment, respectively, were calculated as follows. From Table 1, the grand-mean dim-flash sensitivity (at 520 nm), normalized with respect to the dark current, was $0.33 \text{ photon}^{-1} \mu\text{m}^2$ for rods expressing human or salamander red cone pigment, as compared with $0.52 \text{ photon}^{-1} \mu\text{m}^2$ for wild type, GFP control and human-rhodopsin-expressing *Xenopus* rods combined, giving a ratio of 0.63. Correspondingly, the ratio for the time to peak was 0.81.

Received 19 May; accepted 7 August 2003; doi:10.1038/nature01992.

1. Yau, K. W. Phototransduction mechanism in retinal rods and cones. The Friedenwald Lecture. *Invest. Ophthalmol. Vis. Sci.* **35**, 9–32 (1994).
2. Baylor, D. A. Photoreceptor signals and vision. The Proctor Lecture. *Invest. Ophthalmol. Vis. Sci.* **28**, 34–49 (1987).
3. Okada, T. et al. Circular dichroism of metaiodopsin II and its binding to transducin: a comparative study between meta II intermediates of iodopsin and rhodopsin. *Biochemistry* **33**, 4940–4946 (1994).
4. Imai, H., Imamoto, Y., Yoshizawa, T. & Shichida, Y. Difference in molecular properties between chicken green and rhodopsin as related to the functional difference between cone and rod photoreceptor cells. *Biochemistry* **34**, 10525–10531 (1995).
5. Imai, H., Terakita, A., Tachibana, S., Imamoto, Y., Yoshizawa, T. & Shichida, Y. Photochemical and biochemical properties of chicken blue-sensitive cone visual pigment. *Biochemistry* **36**, 12773–12779 (1997).
6. Starace, D. M. & Knox, B. E. Activation of transducin by a *Xenopus* short wavelength visual pigment. *J. Biol. Chem.* **272**, 1095–1100 (1997).
7. Barlow, H. B. in *Visual Problems of Colour* Vol. 2, 617–630 (HM Stationery Office, London, 1958).
8. Rieke, F. & Baylor, D. A. Origin and functional impact of dark noise in retinal cones. *Neuron* **26**, 181–186 (2000).
9. Witkovsky, P., Levine, J. S., Engbretson, G. A., Hassin, G. & MacNichol, E. F. Jr A microspectrophotometric study of normal and artificial visual pigments in the photoreceptors of *Xenopus laevis*. *Vision Res.* **21**, 867–873 (1981).
10. Palacios, A. G., Srivastava, R. & Goldsmith, T. H. Spectral and polarization sensitivity of photocurrents of amphibian rods in the visible and ultraviolet. *Vis. Neurosci.* **15**, 319–331 (1998).
11. Govardovskii, V. I., Fyhrquist, N., Reuter, T., Kuzmin, D. G. & Donner, K. In search of the visual pigment template. *Vis. Neurosci.* **17**, 509–528 (2000).
12. Moritz, O. L., Tam, B. M., Papermaster, D. S. & Nakayama, T. A functional rhodopsin-green fluorescent protein fusion protein localizes correctly in transgenic *Xenopus laevis* retinal rods and is expressed in a time-dependent pattern. *J. Biol. Chem.* **276**, 28242–28251 (2001).
13. Lyubarsky, A. L., Chen, C.-K., Simon, M. I. & Pugh, E. N. Jr Mice lacking G-protein receptor kinase 1 have profoundly slowed recovery of cone-driven retinal responses. *J. Neurosci.* **20**, 2209–2217 (2000).

14. Baylor, D. A., Matthews, G. & Yau, K.-W. Two components of electrical dark noise in toad retinal rod outer segments. *J. Physiol. (Lond.)* **309**, 591–621 (1980).
15. Rieke, F. & Baylor, D. A. Molecular origin of continuous dark noise in rod photoreceptors. *Biophys. J.* **71**, 2553–2572 (1996).
16. Katz, B. & Miledi, R. The statistical nature of the acetylcholine potential and its molecular components. *J. Physiol. (Lond.)* **224**, 665–699 (1972).
17. Sampath, A. P. & Baylor, D. A. Molecular mechanism of spontaneous pigment activation in retinal cones. *Biophys. J.* **83**, 184–193 (2002).
18. Nakatani, K. & Yau, K.-W. Sodium-dependent calcium extrusion and sensitivity regulation in retinal cones of the salamander. *J. Physiol. (Lond.)* **409**, 525–548 (1989).
19. Ma, J.-X. et al. A visual pigment expressed in both rod and cone photoreceptors. *Neuron* **32**, 451–461 (2001).
20. Miller, J. L. & Korenbrot, J. I. Phototransduction and adaptation in rods, single cones, and twin cones of the striped bass retina: a comparative study. *Vis. Neurosci.* **10**, 653–667 (1993).
21. Tachibana, S., Tsushima, S. & Kawamura, S. Low amplification and fast visual pigment phosphorylation as mechanisms characterizing cone photoreponses. *Proc. Natl Acad. Sci. USA* **98**, 14044–14049 (2001).
22. Miller, J. L., Picones, A. & Korenbrot, J. I. Differences in transduction between rod and cone photoreceptors: an exploration of the role of calcium homeostasis. *Curr. Opin. Neurobiol.* **4**, 488–495 (1994).
23. Kroll, K. L. & Amaya, E. Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signalling requirements during gastrulation. *Development* **122**, 3173–3183 (1996).
24. Wang, Y. et al. A locus control region adjacent to the human red and green visual pigment genes. *Neuron* **9**, 429–440 (1992).
25. Makino, C. L., Groesbeck, M., Lugtenburg, J. & Baylor, D. A. Spectral tuning in salamander visual pigments studied with dihydroretinal chromophores. *Biophys. J.* **77**, 1024–1035 (1999).
26. Merbs, S. L. & Nathans, J. Absorption spectra of human cone pigments. *Nature* **356**, 433–435 (1992).
27. Whitmore, A. V. & Bowmaker, J. K. Seasonal variation in cone sensitivity and short-wave absorbing visual pigments in the rudd *Scardinius erythrophthalmus*. *J. Comp. Physiol. A* **166**, 103–115 (1989).
28. Bowmaker, J. K. & Dartnall, H. J. Visual pigments of rods and cones in a human retina. *J. Physiol. (Lond.)* **298**, 501–511 (1980).
29. Hárosi, F. I. An analysis of two spectral properties of vertebrate visual pigments. *Vision Res.* **34**, 1359–1367 (1994).
30. Hárosi, F. I. Absorption spectra and linear dichroism of some amphibian photoreceptors. *J. Gen. Physiol.* **66**, 357–382 (1975).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Lai, Y. Liang and R. Bruno for help with generating transgenic frogs; J. Nathans for the human rhodopsin and red cone opsin cDNAs and the antibody to red opsin; V. Bhandawat for suggesting the program for computer simulations of random events; and D. A. Baylor, M. E. Burns, S. Kawamura, E. N. Pugh Jr, F. Rieke, J. L. Schnapf, Y. Shichida and members of the Yau laboratory for discussions and comments on the manuscript. This work was supported by the US National Eye Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.-W.Y. (kwyaui@mail.jhmi.edu).

Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels

Xiang Dong Tang¹, Rong Xu¹, Mark F. Reynolds², Maria L. Garcia³, Stefan H. Heinemann⁴ & Toshinori Hoshi¹

¹Department of Physiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

²Department of Chemistry, Saint Joseph's University, 5600 City Avenue, Philadelphia, Pennsylvania 19131, USA

³Department of Ion Channels, Merck Research Laboratories, PO Box 2000, Rahway, New Jersey 07065, USA

⁴Research Unit Molecular and Cellular Biophysics, Medical Faculty of the Friedrich Schiller University Jena, Drackendorfer Strasse 1, D-07747 Jena, Germany

Haem is essential for living organisms, functioning as a crucial element in the redox-sensitive reaction centre in haemproteins¹. During the biogenesis of these proteins, the haem cofactor is typically incorporated enzymatically into the haem pockets of the apo-haemprotein as the functionally indispensable prosthetic

group^{2,3}. A class of ion channel, the large-conductance calcium-dependent Slo1 BK channels, possesses a conserved haem-binding sequence motif. Here we present electrophysiological and structural evidence showing that haem directly regulates cloned human Slo1 channels and wild-type BK channels in rat brain. Both oxidized and reduced haem binds to the hSlo1 channel protein and profoundly inhibits transmembrane K⁺ currents by decreasing the frequency of channel opening. This direct regulation of the BK channel identifies a previously unknown role of haem as an acute signalling molecule.

Database searches showed that Slo1 channels cloned from different species contain the conserved haem-binding amino-acid sequence motif CXXCH (where X is any amino acid) that is typical of c-type cytochromes (Fig. 1a, b). To examine whether haem can acutely regulate the Slo1 channel function, we heterologously expressed human Slo1 (hSlo1) channels and assayed them by using the excised inside-out configuration of the patch-clamp method. In this configuration, membrane patches containing the channels are physically isolated from the rest of the cell and cytoplasmic factors are largely absent.

Application of haemin (Fe³⁺ protoporphyrin IX chloride) at nanomolar concentrations to the intracellular side of the channels consistently and markedly reduced the macroscopic hSlo1 K⁺ current amplitude (Fig. 1c). Mouse Slo1, rat Slo1 and the BK channel, comprising both hSlo1 and the auxiliary subunit β 1, were similarly inhibited (data not shown). The inhibition of the current developed progressively and the time course was accelerated by increasing concentrations of haemin. This potent acute inhibitory action of haemin is notable because, first, the inhibition occurred spontaneously without any exogenously added enzyme or cofactor; and second, the inhibitory effect of haemin significantly outlasted its application duration and was only very slowly reversed by wash, indicative of a very stable physical interaction between haemin and the channel protein. The electrophysiologically measured time course of the hSlo1 current inhibition (Fig. 1c) can be interpreted to reflect spontaneous formation of this putative stable haemin-channel complex.

We found that the inhibition of the Slo1 channel is specific to free haem. The microperoxidase MP-11, in which haem remains incorporated in a small peptide after isolation from horse cytochrome c, did not alter the current amplitude (Fig. 1d, e). FeSO₄ and the haem precursor protoporphyrin IX (haem without iron) did not cause inhibition (Fig. 1d, e). Two other types of metal protoporphyrin IX, cobalt protoporphyrin IX and zinc protoporphyrin IX, decreased the hSlo1 current amplitude, although they were less effective than was haemin at the same concentration (Fig. 1e). Haemin did not show any noticeable effect when applied to the channels from the extracellular side (Fig. 1e). Reduction of the haem iron to the reduced ferrous form (Fe²⁺) by sodium dithionite did not alter the current inhibition efficacy, indicating that the electronic state of the iron centre is not crucial (Fig. 1e).

Several results suggested that the haemin-mediated inhibition of the hSlo1 channel is physiologically relevant. First, the hSlo1 channel was exquisitely sensitive to haemin: the half-maximal inhibitory concentration (IC₅₀) ranged from 45 to 120 nM (Fig. 2a). This concentration is comparable to the dissociation constant (K_d) values of other haemproteins and haem-binding proteins for haem⁴⁻⁶, indicating that the haem-mediated inhibition of the hSlo1 channel comes into play when haem binds to other haem-binding proteins.

Second, native BK channels were similarly inhibited by haemin at the same low concentrations. Haemin decreased the opening frequency of single rat hippocampal BK channels as effectively as that of heterologously expressed hSlo1 channels (Fig. 2b). Notably, the results from single channel proteins also showed that the inhibitory effect of haem (Fig. 1c) is mediated by altering the opening and closing processes to stabilize the nonconducting conformations of the channel, and not by a decrease in the unitary

current amplitude. Even under the inhibitory influence of haemin, individual Slo1 channels retained the ability to open, albeit with very different characteristics.

Third, the haem inhibition persisted at higher, more physiological concentrations of Ca²⁺. The Ca²⁺-dependent activation of Slo1 BK channels, predominantly responding to the local Ca²⁺ concentration, becomes significant at >1 μ M, which is readily achieved in intracellular Ca²⁺ microdomains after cell excitation^{7,8}. With 4.7 μ M Ca²⁺, haemin robustly inhibited the heterologously expressed Slo1 channel currents at physiologically relevant voltages of -60 to 20 mV, at which the plasma membrane potential operates (Fig. 2c, d). Full activation of the channels was achieved with high Ca²⁺ and depolarization to the voltages beyond the physiological range even after haem treatment, further supporting the notion that treatment with haem does not totally eliminate the channel functionality. Similar results with high concentrations of Ca²⁺ were obtained for native rat hippocampal BK channels (data not shown). Together, these characteristics indicate that the inhibitory effect of haem on the Slo1 channel is physiologically relevant.

Inhibition of Slo1 BK channels by haem might be clinically

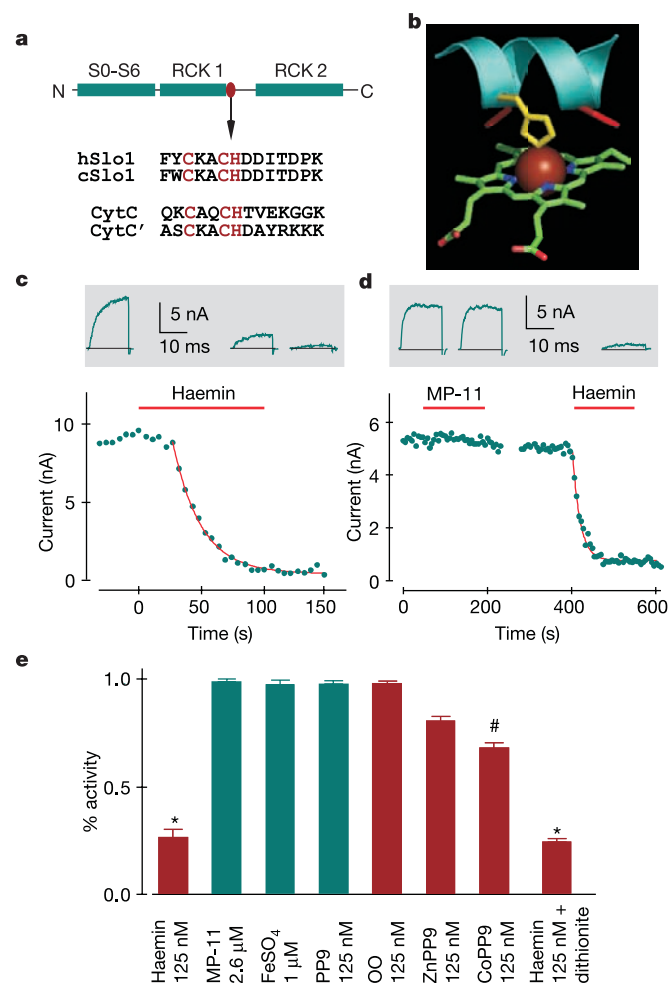


Figure 1 Haemin inhibits the opening of Slo1 channels. **a**, Representation of the haem-binding motif CXXCH in human (GI 7914977) and *Caenorhabditis elegans* (GI 25154297) Slo1, together with horse (GI 117955) and bacterial (GI 117944) cytochromes. **b**, The CKACH segment structure (created with PyMol) of cytochrome c' from *R. palustris* (PDB 1A7V). **c**, Haemin (125 nM) reduces hSlo1 currents at 180 mV with a time constant of 22 s (55 $\pm$ 7 s; n = 6). **d**, MP-11 (2.6 μ M) does not inhibit hSlo1 currents. **e**, Fractional inhibition of hSlo1 currents by haemin and other chemicals. OO, outside-out configuration. Reduction of the haem iron by sodium dithionite (2.5 mM) was confirmed by EPR (not shown). Asterisk, $P < 0.0001$, hash, $P < 0.05$ versus control.

relevant. Acting as the primary cellular Ca^{2+} -excitotoxicity brake, opening of BK channels by pharmacological agents has a cytoprotective role after trauma and ischaemic or hypoxic episodes^{9,10}, two of the known causes of excess haem or haem stress¹¹. The ability of BK channel openers to remove the inhibitory influence of haem may contribute to the cytoprotective function⁹. We demonstrated this by using the benzimidazolone BK channel opener NS1619 and Slo1- β 1 channel complexes in the presence of high Ca^{2+} concentrations to better simulate physiological conditions. NS1619 at 7.5 μM recovered more than $68 \pm 6\%$ of the total current ($n = 3$; Fig. 2e, f), and at a higher concentration (15 μM) essentially eliminated the haem inhibition ($n = 3$; data not shown). Similar results were obtained

using hSlo1 channels without coexpression of β 1 subunits.

We synthesized a 23-residue peptide (hSlo-HBP23) encompassing the putative haem-binding CKACH segment located at residues 612–616 of the channel protein and examined its interaction with haemin by monitoring changes in the ultraviolet–visible and electron paramagnetic resonance (EPR) spectra. Addition of haemin to hSlo-HBP23 in the absence of any cofactor rapidly produced an absorption spectrum characterized by a large peak at about 420 nm (Soret band) and a smaller peak at about 550 nm (α/β -band; Fig. 3a), both of which signify a cytochrome-like coordination of haemin by the peptide. Furthermore, the hSlo-HBP23 absorption spectrum in the presence of haemin closely resembled that of a cytochrome-*c*-derived undecapeptide, MP-11, in which haem remains covalently bound to the amino acid sequence CAQCH.

Hill plot analysis of the hSlo-HBP23 Soret band signal gave a Hill coefficient of about 1 (Fig. 3b), consistent with the idea that one peptide interacts with one haemin. The X-band EPR spectrum of the peptide with haemin showed a distinctive rhombic signal typical of low-spin ferric haem, with two strong axial ligands bound to the haem iron (ref. 12 and Fig. 3c). Comparisons of the EPR parameters with those of other haem proteins suggested that one of the axial ligands might be a histidine nitrogen, consistent with participation of the histidine residue in the sequence CKACH¹³. These spectroscopic measurements suggest that the CKACH segment of the hSlo1 can coordinate haemin in a manner similar to the coordination of haem in cytochromes.

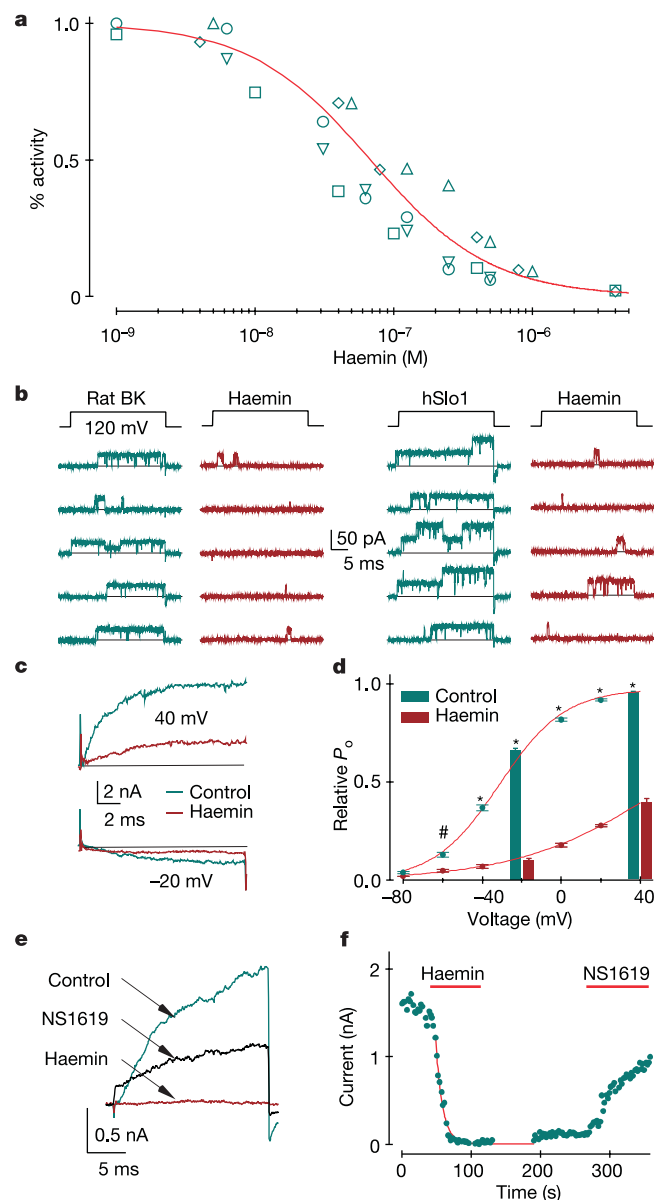


Figure 2 Physiological relevance of haem inhibition. **a**, Concentration dependence tested at 180 mV. $\text{IC}_{50} = 68 \text{ nM}$. **b**, Haemin decreases the opening frequency of rat hippocampal BK (left, $n = 8$) and cloned hSlo1 (right, $n = 7$) channels elicited at 120 mV. **c**, Representative hSlo1 currents at 40 mV and -20 mV recorded at physiological Ca^{2+} (4.7 μM) before and after application of haemin (250 nM). **d**, Haemin reduces hSlo1 currents at physiologically relevant voltages and 4.7 μM Ca^{2+} . Paired *t*-test: Asterisk, $P < 0.01$, hash, $P < 0.05$ versus control. **e**, K^+ currents through hSlo1- β 1 channel complexes recorded with 2 mM K^+ outside and 4.7 μM Ca^{2+} inside at 0 mV were inhibited by haemin (125 nM) and partially recovered by NS1619 (7.5 μM). **f**, Time course of the reversal of haemin-induced hSlo1 inhibition by NS1619.

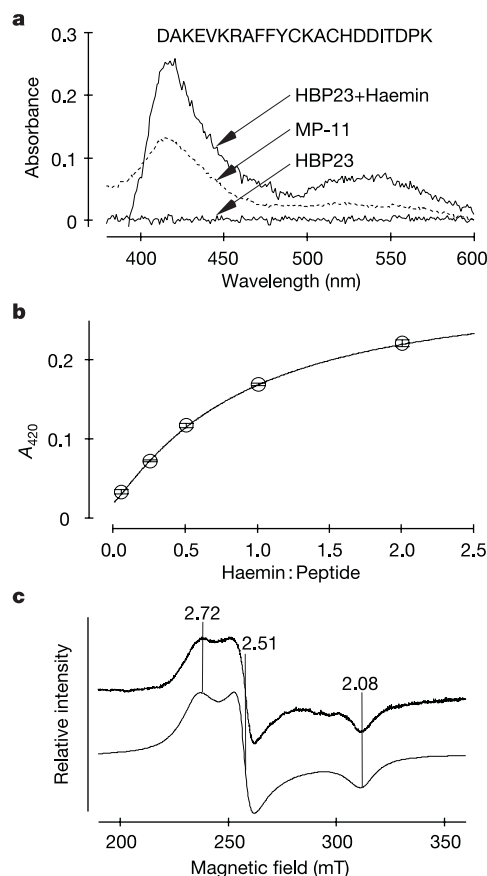


Figure 3 Haemin binding to hSlo-HBP23. **a**, Ultraviolet–visible light spectra of hSlo-HBP23 (40 μM), hSlo-HBP23 (40 μM) with haemin (20 μM), and MP-11 (13 μM plus 2.5 mM dithiothreitol). The difference spectrum is shown for a mixture of hSlo-HBP23 and haemin. **b**, Hill plot analysis of haemin binding by hSlo-HBP23. The absorbance at 420 nm is shown as a function of the haemin:peptide ratio with 20 μM hSlo-HBP23 ($n = 4$). The smooth curve represents a stoichiometry of 1. **c**, EPR spectrum of 1.1 mM hSlo-HBP23 with 1 mM haemin (top) and the best-fit simulation (bottom).

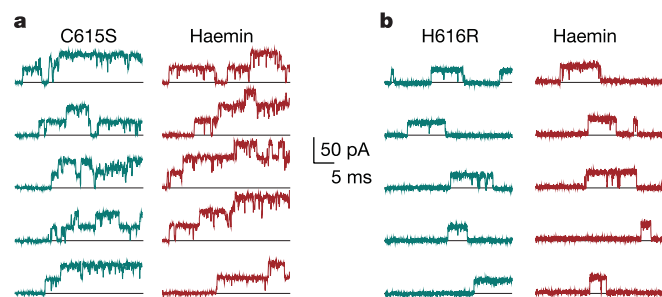


Figure 4 Mutations in the CKACH segment disrupt the haemin sensitivity of the hSlo1 channel. Haemin does not inhibit hSlo1-C615S (**a**) and hSlo1-H616R (**b**) channels. Currents were recorded at 120 mV without added Ca^{2+} . Haemin was applied (250 nM for C615S and 500 nM for H616R) and washed out. Small inhibitory effects observed when haemin was present in the recording chamber were readily reversed by wash.

We tested the involvement of the histidine and cysteine residues in the CKACH segment for haem coordination by site-directed mutagenesis (Fig. 4). The C615S mutation essentially eliminated the sensitivity of the hSlo1 channel to haemin. Similarly, the H616R mutant channel was also insensitive to haemin. These mutagenesis results were consistent with the effects of histidine- and cysteine-modifying reagents on the wild-type channel (Supplementary Fig. 1). Together, these results suggest that the inhibitory effect of haemin on the Slo1 channel gating is mediated by haemin binding to the CKACH segment of the Slo1 channel protein.

The CKACH segment is located between the two putative 'regulator of K^+ conductance' (RCK) domains in the primary amino acid sequence¹⁴. In the *M. thermautotrophicum* K^+ channel MthK, four pairs of RCK domains form a gating ring structure and channel ligands are postulated to control gating by influencing the flexible interface of the RCK domains¹⁴. The haem-binding segment is not present in the MthK structure and how it is positioned relative to the RCK domains is not known. Haem may inhibit the channel opening by influencing the flexible interface.

In summary, we have shown that haem binds to Slo1 BK channels with a high affinity, suggesting that haem may function as an acute cell-signalling molecule. Fast-acting non-genomic modulation of protein function, including that of ion channels, by haem has not been previously documented and the modulation of the Slo1 BK channel reported here represents, to our knowledge, an inaugural case of such a system. After cellular injury, hypoxia and/or stress, haem is liberated from haemoproteins¹¹, leading to an increase in the free haem concentration. The extracellular haem released during hemolysis and trauma can be transported across the plasma membrane¹⁵. In addition, cells may contain other endogenous haem-like molecules that mimic haem action. Inhibition of plasma-membrane Slo1 BK channels is generally expected to exert an excitatory influence on cellular excitability. The existence of Slo1 BK channels in mitochondria, where haem synthesis takes place, has been documented^{10,16,17}. Potential haem regulation of the channels in mitochondria will be particularly significant. □

Methods

Chemicals and stock solutions

Haemin (Fe(III) protoporphyrin IX chloride), Zn(II) protoporphyrin IX, Co(II) protoporphyrin IX and protoporphyrin IX were purchased from Sigma-Aldrich-Fluka. Different batches of haemin produced identical results. Stock solutions (1 mM) of these reagents in 30 mM NaOH were made weekly and kept at -20°C in 1-ml aliquots. The haem concentrations were spectrophotometrically verified¹⁸. Stock solutions (26 mM) of MP-11 (Sigma) were made weekly in water and kept at -20°C . We thawed the solutions immediately before use.

Molecular biology and cell culture

Constructs encoding the Slo1 mutants were generated by polymerase chain reaction with the *pfu* turbo polymerase (Stratagene) and transfected into HEK tsA cells by using

FuGENE 6 (Roche) as described¹⁹. Embryonic rat hippocampal tissues were obtained from BrainBits.

Electrophysiology

The excised configurations of the patch-clamp method were used to record ionic currents in membrane patches isolated from HEK tsA cells and rat hippocampal neurons essentially as described¹⁹. Most of the recordings were done in the inside-out configuration at room temperature ($22-25^\circ\text{C}$). The pipette (extracellular) solution typically contained (in mM) either 120 KCl, 20 KOH, 11 EGTA, 10 HEPES (pH 7.4 with KOH) or 138 NaCl, 2 KCl, 11 EGTA, 10 HEPES (pH 7.4 with NaOH). The bath (intracellular) solution contained (in mM) 120 KCl, 20 KOH, 11 EGTA, 1 MgCl_2 , 10 HEPES (pH 7.4). The high Ca^{2+} solution containing $4.7 \mu\text{M}$ Ca^{2+} was made with $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ and *N*-hydroxyethylenediaminetriacetic acid (HEDTA, Sigma). The free Ca^{2+} concentrations of the solutions were verified with a 97-20 calcium electrode (Thermo Orion). Unless otherwise noted, we carried out the experiments in the virtual absence of Ca^{2+} .

The recording chamber was 200 μl in volume and continuously perfused using a peristaltic pump (Instech) at 1.2 ml min^{-1} . Relative open probability values of hSlo1 channels were estimated by fitting a Boltzmann function to the tail current amplitudes. Because the haem-induced inhibition is long lasting, to measure the inhibition concentration dependence each patch was treated with increasing concentrations. At each concentration, we measured the current amplitude after 4 min.

Peptide synthesis

The hSlo-HBP23 peptide (DAKEVKRAFFYCKACHDDITDPK) was synthesized and purified by HPLC (>95%) by the Protein Chemistry Laboratory facility (University of Pennsylvania). The purified peptide was kept at -20°C and 1-ml solutions, either 1 or 2 mM, were freshly prepared in 100 mM HEPES (pH 7.0) or the electrophysiology bath solution before each use.

Ultraviolet-visible and EPR spectroscopy

Haemin incorporation was achieved by directly mixing haemin and hSlo-HBP23 stock solutions in 100 mM HEPES (pH 7.0). We incubated the sample mixture for 25–30 min before recording the absorption spectra on a Techne spectrophotometer. Difference spectra were obtained by subtracting haemin signals from those of haemin-peptide samples.

To obtain EPR spectra, hSlo-HBP23 (2 mM) was mixed with haemin (2 mM) and hSlo-HBP23 to a final concentration of 1.1 mM hSlo-HBP23 and 1.0 mM haemin. An orange-brown colour change was noticeable when haemin and the peptide were mixed. The samples were then transferred to a quartz EPR tube, frozen in liquid nitrogen at 77 K and tested on a E-109 spectrometer (Varian). Frequencies were determined using a 351D frequency counter (EIP). All experiments were conducted at microwave powers that ensured that resonance saturation did not occur. The EPR simulation was done by Win-EPR SimFonia package v1.0, 1997 (Bruker).

Statistics

Data are given as the mean $\pm$ s.e.m. Analysis of variance (ANOVA) followed by the Scheffe post-hoc test and the paired *t*-test were used as appropriate.

Received 11 June; accepted 26 August 2003; doi:10.1038/nature02003.

- Granick, S. & Beale, S. I. Hemes, chlorophylls, and related compounds: biosynthesis and metabolic regulation. *Adv. Enzymol. Relat. Areas Mol. Biol.* **46**, 33–203 (1978).
- Daltrop, O., Allen, J. W., Willis, A. C. & Ferguson, S. J. *In vitro* formation of a c-type cytochrome. *Proc. Natl Acad. Sci. USA* **99**, 7872–7876 (2002).
- Ponka, P. Cell biology of heme. *Am. J. Med. Sci.* **318**, 241–256 (1999).
- Blackmon, B. J., Dailey, T. A., Lianchun, X. & Dailey, H. A. Characterization of a human and mouse tetrapyrrole-binding protein. *Arch. Biochem. Biophys.* **407**, 196–201 (2002).
- Hirotsu, S. *et al.* Crystal structure of a multifunctional 2-Cys peroxiredoxin heme-binding protein 23 kDa/proliferation-associated gene product. *Proc. Natl Acad. Sci. USA* **96**, 12333–12338 (1999).
- Taketani, S. *et al.* Molecular characterization of a newly identified heme-binding protein induced during differentiation of urine erythroleukemia cells. *J. Biol. Chem.* **273**, 31388–31394 (1998).
- Ricci, A. J., Gray-Keller, M. & Fettiplace, R. Tonotopic variations of calcium signalling in turtle auditory hair cells. *J. Physiol. (Lond.)* **524**, 423–436 (2000).
- ZhuGe, R. *et al.* Dynamics of signaling between Ca^{2+} sparks and Ca^{2+} -activated K^+ channels studied with a novel image-based method for direct intracellular measurement of ryanodine receptor Ca^{2+} current. *J. Gen. Physiol.* **116**, 845–864 (2000).
- Gribkoff, V. K. *et al.* Targeting acute ischemic stroke with a calcium-sensitive opener of maxi-K potassium channels. *Nature Med.* **7**, 471–477 (2001).
- Xu, W. *et al.* Cytoprotective role of Ca^{2+} -activated K^+ channels in the cardiac inner mitochondrial membrane. *Science* **298**, 1029–1033 (2002).
- Dore, S. Decreased activity of the antioxidant heme oxygenase enzyme: implications in ischemia and in Alzheimer's disease. *Free Radic. Biol. Med.* **32**, 1276–1282 (2002).
- More, C. *et al.* EPR spectroscopy: a powerful technique for the structural and functional investigation of metalloproteins. *Biospectroscopy* **5**, S3–S18 (1999).
- Brautigan, D. L. *et al.* Multiple low spin forms of the cytochrome c ferrihemochrome. EPR spectra of various eukaryotic and prokaryotic cytochromes c. *J. Biol. Chem.* **252**, 574–582 (1977).
- Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- Worthington, M. T., Cohn, S. M., Miller, S. K., Luo, R. Q. & Berg, C. L. Characterization of a human plasma membrane heme transporter in intestinal and hepatocyte cell lines. *Am. J. Physiol.* **280**, G1172–G1177 (2001).
- Siemen, D., Loupatatzis, C., Borecky, J., Gulbins, E. & Lang, F. Ca^{2+} -activated K channel of the BK-type in the inner mitochondrial membrane of a human glioma cell line. *Biochem. Biophys. Res. Commun.* **257**, 549–554 (1999).

17. Yermolaieva, O., Tang, X. D., Daggett, H. & Hoshi, T. Calcium-activated mitochondrial K⁺ channel is involved in regulation of mitochondrial membrane potential and permeability transition. *Biophys. J.* **80**, 950A (2001).
18. Shalkai, N., Shviro, Y., Rabizadeh, E. & Kirschner-Zilber, I. Accumulation and drainage of hemin in the red cell membrane. *Biochim. Biophys. Acta* **821**, 355–366 (1985).
19. Tang, X. D. *et al.* Oxidative regulation of large conductance calcium-activated potassium channels. *J. Gen. Physiol.* **117**, 253–274 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. Ciali for critically reading the manuscript; V. Avdonin and R. Wassef for help with constructing the mutants; and P. Angiolillo for use of the EPR spectrometer. This work was supported, in part, by the NIH and the American Heart Association.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to T.H. (hoshi@hoshi.org).

Staphylocoagulase is a prototype for the mechanism of cofactor-induced zymogen activation

Rainer Friedrich^{1*}, Peter Panizzi^{2*}, Pablo Fuentes-Prior¹, Klaus Richter³, Ingrid Verhamme², Patricia J. Anderson², Shun-ichiro Kawabata⁴, Robert Huber¹, Wolfram Bode¹ & Paul E. Bock²

¹Abteilung Strukturforschung, Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany

²Department of Pathology, Vanderbilt University School of Medicine, C3321A Medical Center North, Nashville, Tennessee 37232-2561, USA

³Lehrstuhl Biotechnologie, Technische Universität München, D-85747 Garching, Germany

⁴Department of Biology, Kyushu University, Fukuoka 812-8581, Japan

* These authors contributed equally to this work

Many bacterial pathogens secrete proteins that activate host trypsinogen-like enzyme precursors, most notably the pro-enzymes of the blood coagulation and fibrinolysis systems^{1,2}. *Staphylococcus aureus*, an important human pathogen implicated in sepsis and endocarditis³, secretes the cofactor staphylocoagulase, which activates prothrombin, without the usual proteolytic cleavages, to directly initiate blood clotting^{4,5}. Here we present the 2.2 Å crystal structures of human α -thrombin and prethrombin-2 bound to a fully active staphylocoagulase variant. The cofactor consists of two domains, each with three-helix bundles; this is a novel fold that is distinct from known serine proteinase activators, particularly the streptococcal plasminogen activator streptokinase⁶. The staphylocoagulase fold is conserved in other bacterial plasma-protein-binding factors and extracellular-matrix-binding factors^{7–9}. Kinetic studies confirm the importance of isoleucine 1 and valine 2 at the amino terminus of staphylocoagulase for zymogen activation. In addition to making contacts with the 148 loop and (pro)exosite I of prethrombin-2, staphylocoagulase inserts its N-terminal peptide into the activation pocket of bound prethrombin-2, allosterically inducing functional catalytic machinery. These investigations demonstrate unambiguously the validity of the zymogen-activation mechanism known as ‘molecular sexuality’¹⁰.

During physiological blood coagulation, the zymogen prothrombin is converted to thrombin by cleavage of the single Arg 15-Ile 16 peptide bond (chymotrypsinogen numbering), following the classic mechanism of serine proteinase proteolytic activation. This liberates a new N terminus that inserts into a preformed ‘activation’ or

‘Ile 16’ pocket and triggers the folding of the zymogen ‘activation domain’ (residues 142–152, 186–194 and 216–226) through salt-bridge formation with Asp 194, generating substrate recognition subsites and the oxyanion hole^{11,12}. By contrast, non-proteolytic activation of the zymogens prothrombin and plasminogen by staphylocoagulase (SC) and streptokinase (SK), respectively, is induced by the bacterial cofactors upon formation of tight stoichiometric complexes (see Fig. 1a for the domain organizations of SC and prothrombin). The N-terminal sequences of SC and SK (^{SC}Ile 1-Val-Thr and ^{SK}Ile 1-Ala-Gly) mimic the Ile-Val-Gly/Asn sequences of trypsin and chymotrypsin; this sequence is essentially conserved in most other trypsin-like catalytic domains. Intrusion of these peptides into the activation pockets of the bound prothrombin/plasminogen was proposed to trigger zymogen activation (termed the ‘molecular sexuality’ hypothesis¹⁰). This concept is strongly supported by recent biochemical studies using SK variants^{13,14}, although there is no structural evidence for this mechanism.

Our PSI-BLAST searches revealed strong homologies between SC and four other bacterial proteins that bind to extracellular matrix or plasma proteins (Fig. 1b, and Supplementary Fig. S1); neither SK nor the structurally related staphylokinase share any homology with this family. Characterization of the recombinant SC-like regions of two of these proteins, *S. aureus* von Willebrand factor-binding protein (vWbp)⁷ and *S. agalactiae* NP_687847 (ref. 8), revealed them to be prothrombin activators and showed that vWbp acts allosterically (P.P. and P.E.B., unpublished observations). Therefore, we have grouped SC and its homologues in a novel family of bifunctional bacterial proteins called zymogen-activator and adhesion proteins (ZAAPs).

To characterize the complexes of the prototypical ZAAP, SC, with thrombin and its immediate zymogen precursor, prethrombin-2 (Pre-2), we cloned the active 1–325 fragment⁵ of SC from *S. aureus* Newman 2D. N-terminal sequencing of the fragment showed that the initiating Met residue had not been removed. However, Met-SC(1–325) activated prothrombin (as measured by chromogenic substrate assay), and the complex retained ~80% of the fibrinogen-clotting activity of free α -thrombin. To evaluate the role of the N terminus, a native SC(1–325) fragment was also prepared. The thrombin•Met-SC(1–325) and Pre-2•SC(1–325) crystal structures were determined using phase information obtained by multiple-wavelength anomalous diffraction from a single mercury atom attached to the thrombin-selective thioester-peptide–chloromethyl-ketone inhibitor N^α-[(acetylthio)acetyl]-(D-Phe)-Pro-Arg-CH₂Cl (ATA-FPR-CH₂Cl/ATA-PPACK) and from the correctly positioned thrombin moieties.

The overall structures of both complexes are similar. In particular, the segments that form the activation domain of Pre-2 (which are disordered in the free zymogen¹⁵) are arranged in a thrombin-like manner, that is, the activation domain is correctly folded, with the FPR (D-Phe-Pro-Arg) inhibitor bound in the S1–S4 subsites as observed in the FPR–thrombin structure¹⁶. Both crystals contain two complexes per asymmetric unit, which form a symmetric homodimer (Fig. 2, and Supplementary Fig. S2). Two boomerang-shaped SC moieties stick together through their longer wings, and each prethrombin or thrombin molecule nestles into the elbow of its cognate SC(1–325) molecule. The active sites of the prethrombin and thrombin molecules face a large ‘central cavity’ between the two monomers, and their active site Ser 195 residues are separated by ~73 Å (Fig. 2). A molecular mass of 155 ± 10 kDa was determined for the thrombin•Met-SC(1–325) complex by using sedimentation equilibrium, indicating that the complex dimerizes in solution.

Each SC (activator) molecule consists of two rod-like helical domains connected at an angle of ~110°. The N-terminal domain (D1, residues ^{SC}Ile 1 to ^{SC}Gln 142) essentially consists of three α -helices, which are slightly bent (α_1^D1 , α_3^D1) or considerably twisted (α_2^D1), and slightly wound around each other (Fig. 2). The carboxy-

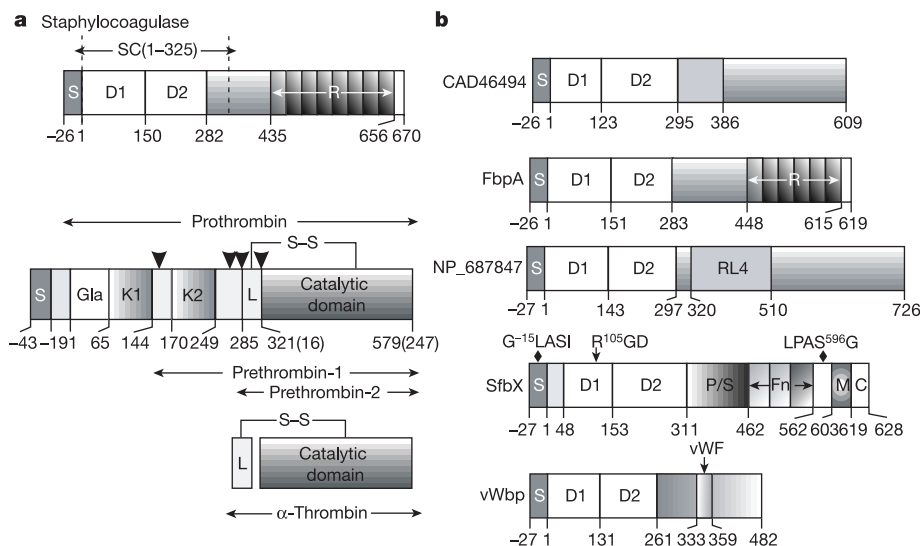


Figure 1 Domain organization of human prothrombin and ZAPs. **a**, Human prothrombin and its cofactor SC. Cleavage sites in the latter are indicated by arrowheads. **b**, SC homologues: CAD46494 (refs 8, 9); FbpA, fibrinogen-binding protein A²²; NP_687847; SfbX, *Streptococcus pyogenes* fibronectin (FN)-binding protein²³; and vWbp⁷. The sequence that directs cell-wall sorting of SfbX is indicated by a diamond; the

RGD triplet is also indicated. Gla, γ -carboxyglutamic acid domain; K, kringle domain; L, light chain; C, cytoplasmic tail; D1/D2, α -helical domains; Fn, FN-binding repeats^{24,25}; M, membrane-spanning peptide; P/S, proline/serine-rich region; R, SC C-terminal repeats; RL4, domain homologous to the large ribosomal subunit L4; S, signal sequence.

terminal half of the longer helix α_2^{D1} and the α_2^{D1} - α_3^{D1} loop protrude out of the helix bundle to form a 'finger'-like structure that is important for dimerization, interlocking with the N-terminal part of helix α_1^{D1} from the opposing monomer. The second domain (D2, residues SC-Thr 150 to SC-Gly 281) starts with a topologically similar bundle formed by helices α_1^{D2} to α_3^{D2} , but exhibits three further α -helices, one of which (α_5^{D2}) is placed on top of the bundle and runs across its axis. The two three-helix bundles are topologically similar (Fig. 2) but only distantly related at the sequence level, suggesting that they arose through a distant gene-duplication event. The domains are likely to fold independently, in agreement with our data showing that prothrombin binding is retained, as is the activator activity of the isolated domains⁵ (see below).

Staphylocoagulase contacts the cognate (pre)thrombin primarily at two surface sites: the relatively flexible 148 loop and the characteristic anion-binding exosite I (Fig. 3). The 148 loop adapts considerably to the SC surface and nestles into a shallow hydrophobic surface groove between SC helices α_1^{D1} and α_2^{D1} , and the α_2^{D1} - α_3^{D1} connector (Fig. 3a). The Trp 148 side chain of thrombin, which is involved in the majority of contacts, moves ~ 10 Å relative

to its position in FPR-thrombin¹⁶. Central to the second interface is the side chain of a thrombin residue (Tyr 76) that is critical for substrate/cofactor recognition. Tyr 76 occupies a hydrophobic cavity lined by side chains from both moieties (Fig. 3b). The adjacent guanidyl groups of thrombin residues Arg 73, Arg 75 and Arg 77A primarily contact acidic residues of the cofactor (for example, SC-Glu 46, SC-Glu 213 and SC-Asp 217). Thus, SC differs from SK and staphylokinase not only in its overall structure but also in its contact sites; SK and staphylokinase¹⁷ consist mainly of antiparallel β -sheets and have a different arrangement around the plasminogen catalytic domain (see Supplementary Movie S3).

To determine whether the low-affinity precursor form of exosite I on prothrombin (proexosite I¹⁸) is engaged by SC, we used 5-(carboxy)fluorescein-labelled hirudin(54-65) ([5F]Hir(54-65)(SO₃⁻)) as an exosite-I-specific probe¹⁸. Fluorescence titrations of the labelled peptide with prothrombin or thrombin showed that Met-SC(1-325) is a tight, competitive inhibitor of [5F]Hir(54-65)(SO₃⁻) binding to (pro)exosite I, with stoichiometries of 0.8 ± 0.1 mol Met-SC(1-325) per mol prothrombin (Fig. 3c) and 0.94 ± 0.03 mol Met-SC(1-325) per mol thrombin (Fig. 3d).

Further experiments dissected the contribution of the individual SC domains to prothrombin activation. Fragment SC(1-234) exhibited a marginal decrease (a $\sim$ fivefold reduction) in activity compared with SC(1-325), whereas domain D1 (SC(1-146)) by itself activated prothrombin fully, although it had a $\sim 3,000$ -fold lower affinity (dissociation constant, $K_D = 780 \pm 150$ nM) for prothrombin (Fig. 4a). By contrast, SC(147-234) did not activate prothrombin at concentrations up to 43 μ M.

To assess quantitatively the role of the natural N terminus of SC in zymogen activation, we compared the activity of native SC(1-325) with those of Met-SC(1-325) and other N-terminally truncated variants. SC(1-325) activated prothrombin stoichiometrically and had an apparent K_D of 0.3 ± 0.2 nM, representing a 60-fold higher potency than the Met-SC(1-325) analogue ($K_D = 17 \pm 2$ nM) (Fig. 4b). SC(2-325) activated prothrombin with a $\sim$ sixfold lower affinity compared with native SC(1-325), whereas, by contrast, SC(3-325) exhibited $<2\%$ prothrombin-activator activity. We conclude that domain D1 and particularly the N-terminal SC dipeptide are essential and sufficient for non-proteolytic activation

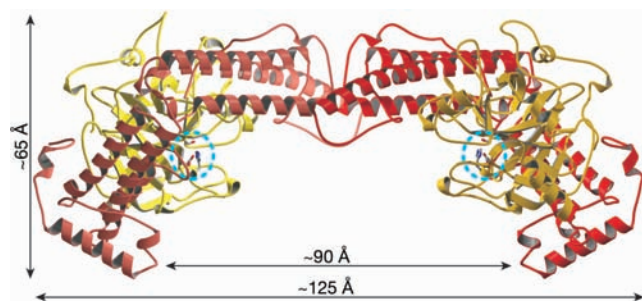


Figure 2 Structure of the human α -thrombin•Met-SC(1-325) complex. Ribbon plot showing the symmetric dimer in the asymmetric unit. Thrombin is shown in yellow and gold; SC is shown in red and salmon. The side chains of thrombin's active-site residues are shown with all non-hydrogen atoms and are circled in blue. Both monomers are related by an exact but crystallographically 'local' two-fold rotation axis.

of prothrombin, whereas domain D2 is needed for high-affinity binding of prothrombin through proexosite I.

The crystal structure of the Pre-2•SC(1–325) complex verifies the role of the N-terminal SC residues in prothrombin activation (see Supplementary Movie S4). Unlike the thrombin complex, the hexapeptide ^{SC}Ile 1 to ^{SC}Tyr 6 is fully defined by electron density (Fig. 4d). The first four residues of this peptide occupy a position similar to that of the endogenous Ile 16 to Gly 19 peptide. The free ^{SC}Ile 1 α -ammonium group reaches into the negatively charged pocket created by the carboxylate of Asp 194 and neighbouring carbonyls of Pre-2, forming a buried salt bridge (Fig. 4c). The ^{SC}Ile 1–^{SC}Tyr 6 ‘activation peptide’ is also stabilized through main-chain to main-chain hydrogen bonds of ^{SC}Val 2 to the specificity-determining residue Asp 189 and additional van der Waals interactions. In agreement with previous kinetic studies, modelling experiments showed that the N-terminal peptide could be either extended or shortened by one residue and still reach the Ile 16 pocket, whereas a form lacking the first two residues would be too short to access the activation pocket. The Pre-2•SC(1–325) structure and the mutation data thus provide direct proof of the

N-terminal insertion, as predicted in the molecular sexuality hypothesis. The observation that zymogen activation can tolerate significant variations in both length and sequence of the SC N terminus reveals an unexpected degree of promiscuity. The finding that plasminogen activation by SK also requires the N terminus^{13,14}, despite the fact that the domain organization, structure and binding topology of SK⁶ are completely unrelated to SC, supports the general validity of the cofactor-induced activation mechanism.

In summary, the crystal structures of SC(1–325) bound to α -thrombin and its precursor Pre-2 provide clear evidence of the molecular sexuality mechanism for non-proteolytic zymogen activation. The previously uncharacterized SC fold is shared by several other bacterial proteins, which define a new class of conformational activators, ZAAPs. SC uncouples thrombin activities in blood coagulation and fibrinolysis from their physiological regulation, resulting in fibrin deposition in acute bacterial endocarditis. Importantly, SC and vWbp binding to fibrinogen and von Willebrand factor may trigger prothrombin activation at sites of platelet aggregation and thrombosis, which will expand the bacterial fibrin–platelet vegetation on damaged endothelium. Our results

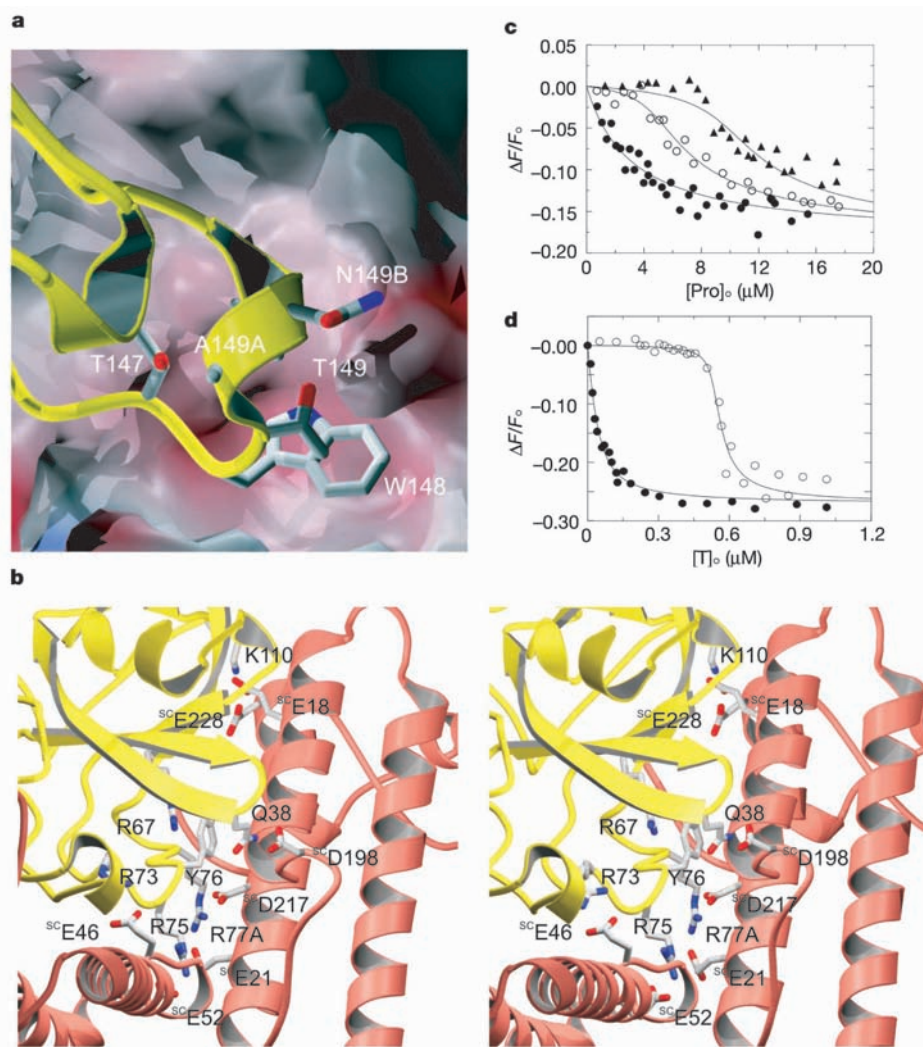


Figure 3 Thrombin–staphylocoagulase interface. **a**, The 148-binding pocket. SC is shown as a solid-surface model coloured according to its electrostatic surface potential (red, negative; blue, positive), and thrombin is represented as a ribbon. **b**, The exosite I–D2 interaction. Stereo plot showing the SC (red) and the thrombin backbone (yellow) together with a few interacting residues. **c–d**, Effect of Met-SC(1–325) on [5F]Hir(54–65)(SO₃[−]) binding to exosite I of prothrombin. **c**, Fractional changes in fluorescence ($\Delta F/F_0$) of

50 nM [5F]Hir(54–65)(SO₃[−]) as a function of total native prothrombin concentration ($[Pro]_0$) in the absence (filled circles) or presence of 4 μM (open circles) or 8 μM (filled triangles) Met-SC(1–325). **d**, Titrations of 36 nM [5F]Hir(54–65)(SO₃[−]) with thrombin (T) in the absence (filled circles) or presence (open circles) of 0.5 μM Met-SC(1–325). Lines represent nonlinear least-squares fits by a model for competitive binding using the parameters given in the text.

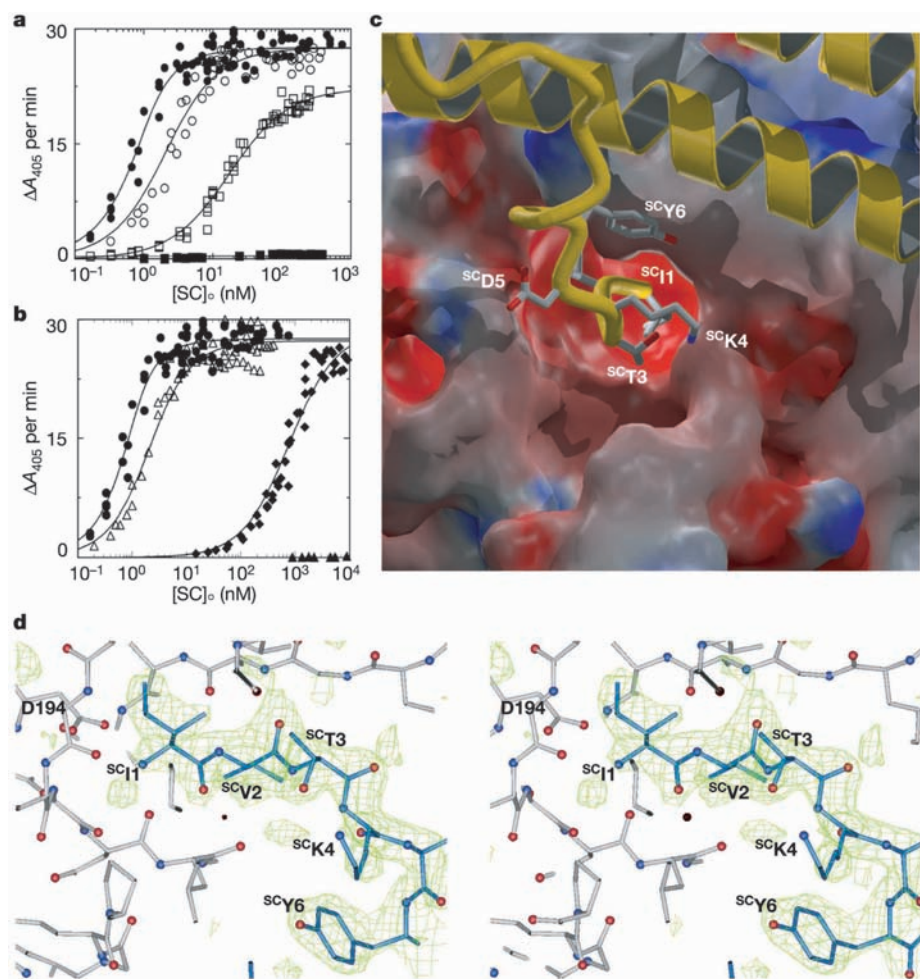


Figure 4 Contribution of the staphylocoagulase N-terminal peptide to zymogen activation. **a**, Titrations of the increase in the rate of chromogenic substrate hydrolysis (ΔA_{405} per min) of 1 nM prothrombin with SC(1–325) (filled circles), Met-SC(1–325) (open squares), SC(2–325) (open circles), and SC(3–325) (filled squares). The lines represent the least-squares fit using the binding equation with the K_D values given in the text. **b**, Activity titrations of prothrombin with SC(1–325) (filled circles), SC(1–234) (open triangles),

SC(1–146) (filled diamonds) and SC(147–325) (filled triangles). **c**, Binding of the N-terminal SC11–SC6 peptide to Pre-2. The SC backbone is shown in yellow, its N-terminal side chains have coloured atoms, and the bound zymogen is coloured according to its electrostatic surface potential. **d**, Detail of the initial $F_o - F_c$ electron-density maps of the Pre-2•SC complex (green; contoured at 1 σ). The fully defined SC11–SC6 peptide is highlighted in blue.

Table 1 Statistics of data collection and refinement

Crystal	FPR-thrombin-Met-SC(1-325)			FPR-Pre-2-SC(1-325)		
Space group	C2			C2		
Molecules in asymmetric unit	2			2		
Data collection						
Resolution (Å)	2.2			2.1		
a (Å)	180.17			181.16		
b (Å)	102.00			102.34		
c (Å)	135.30			135.54		
β (°)	130.08			129.91		
Wavelength (Å)	1.0085	1.0050	0.9200	1.0092	1.0000	0.9500
R _{sym}	0.048	0.044	0.042	0.048	0.039	0.042
R _{sym} (outermost shell, %)	0.269	0.232	0.236	0.269	0.260	0.243
Unique reflections	95,100	95,248	95,061	110,873	110,836	110,669
Completeness (%)	97.5	97.4	91.1	90.4	91.1	94.3
Completeness (outermost shell, %)	96.4	96.5	35.4	74.7	83.4	90.6
Refinement						
Resolution (Å)	2.2			2.2		
Test set size (%)	5.0			5.0		
Reflections used for refinement	90,671			87,632		
Final R _{free} (%)	24.9			25.3		
Final R (%)	20.9			21.0		
r.m.s.d. bonds (Å)	0.011			0.011		
r.m.s.d. angles (°)	1.60			1.60		
Mean B value (Å²)	37.0			39.2		
Protein atoms	9,370			9,418		
Heterogen atoms	42			42		
Solvent atoms	748			793		

might help to develop new therapies based on the inhibition of pathological processes invoked by these bacterial proteins. □

Methods

Protein expression, purification and characterization

Human prothrombin and factor X were purified and activated as previously described¹⁹. SC(1–325) was cloned by polymerase chain reaction from the genomic DNA of *S. aureus* Newman D2, strain Tager 104 (ref. 20), and expressed in *Escherichia coli* BL21(DE3) plysS cells. Met-SC(1–325) was purified by affinity chromatography on prothrombin–Affigel 10, followed by gel filtration on Superdex 200 HR 10. Native SC(1–325) was expressed as a His₆-tagged fusion protein with a factor Xa cleavage site encoded after the initiating Met. The fusion protein was expressed as described for Met-SC(1–325), purified by chromatography on nickel–iminodiacetic acid Sepharose, cleaved with factor Xa, and separated from the tag. SC(2–325) and SC(3–325) were prepared similarly by successive deletion. C-terminally truncated domain mutants were generated by inserting stop codons using QuickChange site-directed mutagenesis. SC(147–325) was prepared following a procedure similar to that for SC(1–325). The identities of SC N-terminal mutants were verified by N-terminal protein sequencing.

Activity and fluorescence studies

Prothrombin–SC activity titrations were measured from the initial rate of 100 μM H-D-Phe-Pip-Arg-*p*-nitroanilide hydrolysis at 405 nm in 50 mM HEPES, 110 mM NaCl, 5 mM CaCl₂, 0.1 mg ml^{−1} soybean trypsin inhibitor, 1 mg ml^{−1} BSA, pH 7.4 at 25 °C. Prothrombin and SC variants were preincubated for 20 min before addition of the substrate. Titrations were fitted by the quadratic binding equation to obtain the maximum activity and apparent *K*_D. Fluorescence was measured with an SLM 8100 fluorometer following previously described methods^{17,21}. All experiments were performed in the above-described buffer without soybean trypsin inhibitor and at 25 °C. Binding of [5F]Hir(54–65)(SO₃[−]) was analysed by fitting of the quadratic binding equation to obtain the maximum fluorescence change, *K*_D and stoichiometric factor (*n*). The effect of Met-SC(1–325) on [5F]Hir(54–65)(SO₃[−]) binding was measured in prothrombin or thrombin titrations of mixtures of labelled peptide and Met-SC(1–325). The data were fitted using the cubic equation for tight competitive binding to determine *K*_D and *n*. Error estimates represent ±2 s.d.

Crystallization and structure determination

Thrombin was inactivated with a sevenfold excess of N^α-[(acetylthio)acetyl]-(D-Phe)-Pro-Arg-CH₂Cl as described previously¹⁹, and reacted with a 20-fold molar excess of *p*-hydroxymercuribenzoate in the presence of 0.1 M NH₂OH for 3 h at pH 7.0 and 25 °C to generate and label the free thiol. Crystals of the α-thrombin•Met-SC(1–325) and the Pre-2•SC(1–325) complex were obtained by hanging-drop vapour diffusion from protein solutions containing 100 mM imidazole (pH 7.5), 200 mM sodium formate, 12% (w/v) PEG 4000, using microseeding; crystals contained two (pre)thrombin•SC complexes per asymmetric unit. Three-wavelength anomalous-diffraction experiments were conducted at the beam line BW6 (Deutsches Elektronen Synchrotron, Hamburg), using crystals flash-cooled to 100 K in the presence of 15% butane diol. Bijvoet pairs to 2.20 Å resolution were collected in two continuous 100° sweeps using a MAR-CCD detector (Table 1). All reflection data were processed using DENZO/SCALEPACK (http://krzys.med.virginia.edu/CrystUVA/HKL_page.html). Data sets collected at the Hg LIII absorption edge (λ₁) and the inflection point (λ₂) were scaled to the remote (λ₀) data set using CNS (Crystallography and NMR System; <http://cns.csb.yale.edu/>). A Patterson search with AMoRe (<http://www.ccp4.ac.uk/dist/html/amore.html>) using the coordinates of FPR-α-thrombin (1PPB¹⁶) yielded the positions of the thrombin molecules. The electron-density map calculated with the derived phases showed unambiguous density for the modified FPR moieties, including the attached mercury atoms. The mercury positions were refined, and anomalous experimental phases at 2.4 Å were calculated using CNS, yielding a final figure of merit of 0.875 after iterative solvent flattening. Density modification and non-crystallographic symmetry averaging improved the combined experimental and model phases. The resulting electron density allowed us (using MAIN; <http://www.bmb.ijs.si/doc/>) to build a protein model that comprised complete α-thrombin chains from Glu 1C to Tyr 14H and Ile 16 to Gly 246, and SC residues ^{SC}Ser 7 to ^{SC}Gly 281. The C-terminal segments (^{SC}Glu 282–^{SC}Gln 325) were not defined by electron density, which suggests enhanced flexibility, as predicted by secondary-structure prediction methods (<http://cubic.bioc.columbia.edu/predictprotein/>; <http://www.pondr.com/>). The α-thrombin•Met-SC(1–325) model was refined with CNS against the λ₀ data set. The Pre-2•SC(1–325) structure was solved using a similar approach. The main-chain angles of all non-glycine residues in both structures fall into allowed or additionally allowed regions. The final refinement parameters are given in Table 1.

Database searches

Searches for sequence and structural homologues were performed with PSI-BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and DALI (<http://www.ebi.ac.uk/dali/>), respectively.

Received 13 May; accepted 25 July 2003; doi:10.1038/nature01962.

- Boyle, M. D. & Lottenberg, R. Plasminogen activation by invasive human pathogens. *Thromb. Haemost.* **77**, 1–10 (1997).
- Lähteenmäki, K., Kuusela, P. & Korhonen, T. K. Bacterial plasminogen activators and receptors. *FEMS Microbiol. Rev.* **25**, 531–552 (2001).
- Mylonakis, E. & Calderwood, S. B. Medical progress: Infective endocarditis in adults. *N. Engl. J. Med.* **345**, 1318–1330 (2001).
- Hemker, H. C., Bas, B. M. & Muller, A. D. Activation of a pro-enzyme by a stoichiometric reaction with another protein. The reaction between prothrombin and staphylocoagulase. *Biochim. Biophys. Acta* **379**, 180–188 (1975).
- Kawabata, S. *et al.* Structure and function relationship of staphylocoagulase. *J. Prot. Chem.* **6**, 17–32 (1987).
- Wang, X., Lin, X., Loy, J. A., Tang, J. & Zhang, X. C. Crystal structure of the catalytic domain of human plasmin complexed with streptokinase. *Science* **281**, 1662–1665 (1998).
- Bjerketorp, J. *et al.* A novel von Willebrand factor binding protein expressed by *Staphylococcus aureus*. *Microbiology* **148**, 2037–2044 (2002).
- Tettelin, H. *et al.* Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc. Natl Acad. Sci. USA* **99**, 12391–12396 (2002).
- Glaser, P. *et al.* Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol. Microbiol.* **45**, 1499–1513 (2002).
- Bode, W. & Huber, R. Induction of the bovine trypsinogen–trypsin transition by peptides sequentially similar to the N-terminus of trypsin. *FEBS Lett.* **68**, 231–236 (1976).
- Huber, R. & Bode, W. Structural basis of the activation and action of trypsin. *Acc. Chem. Res.* **11**, 114–122 (1978).
- Khan, A. R. & James, M. N. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **7**, 815–836 (1998).
- Boxrud, P. D., Verhamme, I. M., Fay, W. P. & Bock, P. E. Streptokinase triggers conformational activation of plasminogen through specific interactions of the amino-terminal sequence and stabilizes the active zymogen conformation. *J. Biol. Chem.* **276**, 26084–26089 (2001).
- Wang, S., Reed, G. L. & Hedstrom, L. Deletion of Ile1 changes the mechanism of streptokinase: evidence for the molecular sexuality hypothesis. *Biochemistry* **38**, 5232–5240 (1999).
- Vijayalakshmi, J., Padmanabhan, K. P., Mann, K. G. & Tulinsky, A. The isomorphous structures of prethrombin2, hirugen-, and PPACK-thrombin: changes accompanying activation and exosite binding to thrombin. *Protein Sci.* **3**, 2254–2271 (1994).
- Bode, W. *et al.* The refined 1.9 Å crystal structure of human α-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trip insertion segment. *EMBO J.* **8**, 3467–3475 (1989).
- Parry, M. A. *et al.* The ternary microplasmin–staphylokinase–microplasmin complex is a proteinase–cofactor–substrate complex in action. *Nature Struct. Biol.* **5**, 917–923 (1998).
- Anderson, P. J., Nesset, A., Dharmawardana, K. R. & Bock, P. E. Characterization of proexosite I on prothrombin. *J. Biol. Chem.* **275**, 16428–16434 (2000).
- Bock, P. E. Active-site-selective labeling of blood coagulation proteinases with fluorescence probes by the use of thioester peptide chloromethyl ketones II. Properties of thrombin derivatives as reporters of prothrombin fragment 2 binding and specificity of the labeling approach for other proteinases. *J. Biol. Chem.* **267**, 14974–14981 (1992).
- Tager, M. & Drummond, M. C. Staphylocoagulase. *Ann. NY Acad. Sci.* **128**, 92–111 (1965).
- Bock, P. E., Olson, S. T. & Björk, I. Inactivation of thrombin by antithrombin is accompanied by inactivation of regulatory exosite I. *J. Biol. Chem.* **272**, 19837–19845 (1997).
- Cheung, A. I., Projan, S. J., Edelstein, R. E. & Fischetti, V. A. Cloning, expression, and nucleotide sequence of a *Staphylococcus aureus* gene (fbpA) encoding a fibrinogen-binding protein. *Infect. Immun.* **63**, 1914–1920 (1995).
- Jeng, A. *et al.* Molecular genetic analysis of a group A *Streptococcus* operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. *J. Bacteriol.* **185**, 1208–1217 (2003).
- Joh, D., Wann, E. R., Kreikemeyer, B., Speziale, P. & Höök, M. Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells. *Matrix Biol.* **18**, 211–223 (1999).
- Schwarz-Linek, U. *et al.* Pathogenic bacteria attach to human fibronectin through a tandem β-zipper. *Nature* **423**, 177–181 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to Diagnostica Stago for the *S. aureus* strain, to R. Mentle for sequencing, to G. Bourenkov for help during data collection, and to S. Iwanaga for initiating the project. We thank E. Consacro, H. Kroh and S. Stuart for excellent technical assistance. This work was supported by a National Institutes of Health (NIH) grant to P.E.B., as well as funding from SFB469, the EU project SPINE and the Fonds der Chemischen Industrie. P.P. and P.J.A. were supported in part by NIH training grants.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.B. (bode@biochem.mpg.de) or P.E.B. (paul.bock@vanderbilt.edu). The sequence of SC Newman D2 strain Tager 104 has been deposited in GenBank under accession code AY225090. The atomic coordinates of human α-thrombin•Met-SC(1–325) and of human Pre-2•SC(1–325) have been deposited in the Protein Data Bank under accession codes 1NU7 and 1NU9, respectively.

Codon optimization to PCR

Recent introductions include an array of heart disease genes.

A² reader

Beckman Coulter www.beckmancoulter.com

Fluorescence-based CCD microplate reader

The A² is designed for automatic quantification of hundreds of protein analytes on the A² microarray plate as part of Beckman Coulter's A² Microarray System for multiplexed immunoassays. The A² system allows users to detect and quantify protein–protein interactions, protein immunoassays, DNA–protein interactions, enzyme reactions and other



A square deal: the A² microplate reader.

protein assays. The reader, about the size of a laptop computer, can read a plate and process microarray data in under five minutes. Built-in software aligns a grid around each of the spots on the microarray plates and measures the fluorescent intensity. The A² reader can also be integrated with Biomek automated liquid-handling platforms.

Cardiovascular disease gene array II

Superarray Bioscience www.superarray.com

Vital information

This new product in the GEArray Q Series is designed to analyse the relative expression of 96 atherosclerosis-related genes. The human genes on this array are important for the onset, progression and prevention of atherosclerosis. Through simple side-by-side hybridization, investigators can simultaneously determine the expression profile of these endothelial cell genes using experimental RNA samples, the arrays provided in the kit, and a SuperArray Probe Synthesis kit. An array containing the equivalent mouse genes will also be available.

Advalytix ArrayBooster

GRI Molecular Biology www.gri.co.uk

Agitation beats diffusion

Most microarray substrates are conventional microscope slides with up to several thousand spots of oligonucleotides or cDNA probes with known identity covering the slide in a

checkerboard pattern. The sample is usually sandwiched between the DNA microarray and a cover slip, forming a reaction chamber, and diffusion is the only mechanism by which the DNA strands move within the capillary gap. The Advalytix ArrayBooster uses agitation based on surface acoustic waves, rather than diffusion, to mix the sample during incubation. This allows the system to reach equilibrium in an overnight hybridization, increases the signal-to-noise ratio, improves slide-to-slide reproducibility and enhances homogeneity across the array.

Platinum PCR SuperMix High Fidelity

Invitrogen www.invitrogen.com

Double advantage

The Platinum PCR SuperMix High Fidelity is described as combining the robustness, fidelity and yields of Platinum Taq DNA Polymerase High Fidelity with the time saving, simplicity and high productivity of a PCR supermix. The enzyme increases PCR fidelity approximately sixfold over Taq polymerase alone. The format is pre-mixed and needs only to be added to template and primers to begin PCR.

Codon optimization service

GenScript www.genscript.com

A boost for protein expression

Codon preferences can vary widely between species. To enhance the expression level of a foreign protein in a given system (*E. coli*, yeast, insect or mammalian cell), it is important to adjust the codon frequency of the foreign protein to match the host expression system. An algorithm developed by GenScript optimizes sequences for protein expression based on the user's codon table, or on a publicly available codon usage database. The algorithm converts an amino acid sequence into a DNA sequence with overall codon usage similar to a specified organism, and optimizes RNA secondary structure, GC content and repetitive codons.

Low-volume UV microplates

Corning Life Sciences www.corning.com

Do it by halves

Corning's half-area, 96-well UV microplate for quantification of DNA, RNA and proteins features a novel well height that creates a sample volume of 175 µl. This corresponds directly with a 1 cm path length for calibration using UV absorbance, at the same time offering reductions in reagent costs and the use of limited samples. Corning claims that researchers will save significant amounts of time by using UV microplates rather than quartz cuvettes.

The disposable microplates eliminate washing and recycling, and also remove the risk of damaging fragile and expensive quartz cuvettes. A single half-area microplate is sufficient for all steps, from calibration to direct measurements of nucleotide or protein concentrations. The plates are made of high-transmittance polymer that has low background for consistent detection of samples at 260 nm and 280 nm. The UV-transparent bottom is moulded directly onto the plate, creating an integral seal that prevents leaks and cross-contamination.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
**Netherlands/ Italy/
Iberia/ Belgium:**
Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)
France/ Switzerland:
Amelie Pequignot (4974)
Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

Physics temps

Academic physics jobs in the United States are in transition, with a growing emphasis on part-time or temporary appointments, a continuing decrease in faculty members born in the United States and an economic crunch that could result in fewer permanent positions, according to a recently released survey by the American Institute of Physics (www.AIP.org).

Faculty at US physics departments grew by 5% from 2000 to 2002 — but much of that growth is due to an increase in temporary and part-time instructors. This reflects a softening in the academic market that has already hit the humanities and is now showing signs of spreading to all the sciences. This trend may worsen as state budgets, which fund many of the large public universities, tighten as a result of recent tough economic times, according to the AIP, which conducts the physics survey every two years.

There are still jobs, however. The number of tenure-track jobs and the number of tenured physics faculty that have retired have remained relatively unchanged since 2000. The wild card is the economy. If state budgetary shortfalls leave public universities in the lurch, which is often the case, universities may opt to let vacant positions go unfilled or use temporary help to make up the difference.

Getting those positions may become tougher — unless fewer foreign physicists flow into the United States as a result of visa restrictions. With most of the growth being in part-time positions, and with retirement levels relatively flat, there will be greater competition for tenure-track positions. It may still be possible to climb the tenure-track ladder in physics, but the starting rung will often be lower, in the form of a temporary or a part-time position.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS AND STUDENTS

Postdocs cross disciplines p542

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Going multidisciplinary

A varied training path is trickier to navigate than traditional specialized routes — but can be more rewarding, says Myrna Watanabe

Although most graduate students and postdocs still stick to a single research field, an increasing number are looking to combine a number of different disciplines. Young researchers seek broader horizons for many reasons. Some see no future in the field they are in. Others realize that they need to add another layer of skills. Yet others enjoy the cross-pollination in a multidisciplinary team. And some have been roaming across disciplines from their undergraduate years and cannot imagine doing research any other way.

Government initiatives in the United States and Europe are helping by pushing for 'one-stop shopping' for cross-disciplinary training. But there is no single starting point. Some students start thinking multidisciplinary as undergraduates, some seek more breadth as graduate students and some new PhDs feel constrained by their specialities, and so take multidisciplinary fellowships. There is a similar diversity of organization. Multidisciplinary programmes can be run by a single department, or can be interdepartmental or involve several schools.

The Vertical Integration of Research and Education (VIGRE) programme in the Division of Mathematical Sciences of the US National Science Foundation (NSF) trains undergraduates, graduate students and postdoctoral fellows. Programme director Richard Millman notes that the shift towards broader multidisciplinary training is relatively new. In the past, one collaborator would handle, say, the mathematics, and the other the biology. Now, "two people need to be experts in both areas before moving on," Millman says.

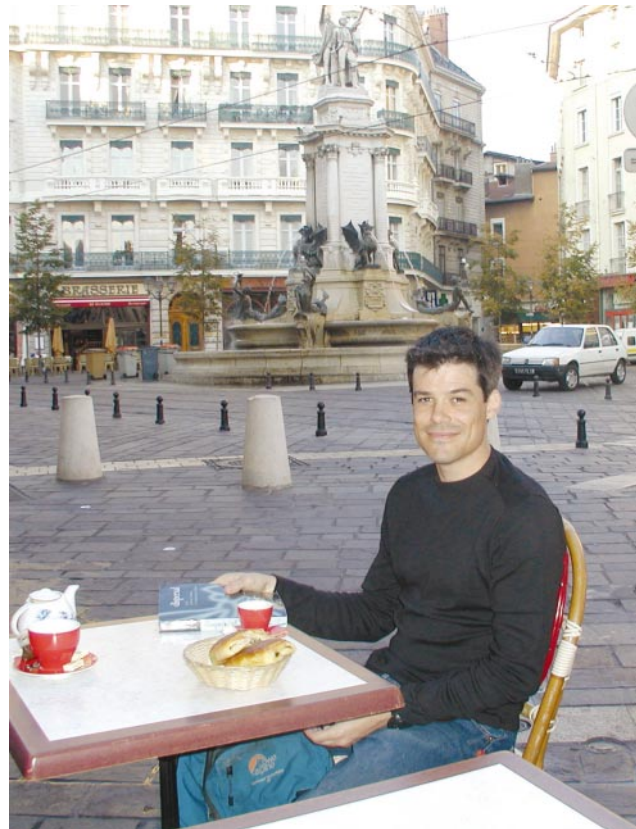
However, Charles DeLisi, Metcalf Professor of Science and Engineering at Boston University, and former dean of engineering, cautions that mixing disciplines is not for everyone. People who prefer to be highly specialized should take a more traditional route to their scientific career. But "people who like to see relationships and enjoy breadth" would benefit from a multidisciplinary route, says DeLisi.

GETTING A GRANT

There are now more programmes than ever before to help graduate students and postdocs find diversity and breadth in their training. There is no comprehensive listing (see Web links for a selection), so postdocs who want a multidisciplinary placement have to think creatively. You may find one through an ad in a scientific journal or a professional organization's job listings. Other ways are through faculty recommendation, knowing a potential postdoctoral sponsor, or Internet searches for labs active in a



Richard Millman: giving grants for broader training.



Conservation biologist David Tallmon wonders whether his interdisciplinary skills will be valued more outside academia.

particular field. Tracking published papers of interest and then contacting the head of the lab has been successful for some. Prolific labs may get a surfeit of applicants. "I probably get at least an application per day," says Yury Gogotsi, the principal investigator on an NSF Integrative Graduate Education and Research Training (IGERT) grant on nanoscale engineering at Drexel University, Philadelphia, Pennsylvania. His doctoral student, chemist Nevin Naguib, whose thesis is on microfluidics of materials filled into carbon nanotubes, has begun her postdoctoral search by looking on the web. She wants to stay in nanotechnology and hopes to be able to mix chemistry and materials science.

Because multidisciplinary programmes are relatively new, postdocs and graduate students also have to be creative in finding funding. Many have used the IGERT programme, which gives grants for interdisciplinary graduate training to institutions. Unfortunately, IGERT grants only aid graduate education and do not cover postdocs, although many investigators funded under IGERT find funds for postdocs from other sources.

In the United States, the NSF and the National Institutes of Health (NIH) also make grants to individual postdocs for

multidisciplinary work in the United States or abroad. But some grants go unused because few appropriately trained potential postdocs are aware of them. "We get very few applications," says Milton Hernandez, director of the Office of Special Populations and Research Training for the National Institute of Allergy and Infectious Diseases at the NIH. His programme gives postdoctoral awards for PhDs with training in mathematics, engineering and computer sciences who want to work in biological research related to human health.



Going nanoscale: chemist Nevin Naguib wants to mix chemistry and materials science.

mathematics at Rice University, in Houston, Texas. The programme, which combines research with teaching at both the graduate and undergraduate levels, is good preparation for a career in academia, he says.

Chemical engineer Andrew Metters, whose NSF-funded postdoc was on bioengineering materials for nervous system repair at the Swiss Federal Institute of Technology (ETH) in Zurich, has already entered on a career as an assistant professor in the department of chemical engineering at Clemson University, South Carolina. Not only is he continuing his

PICKING A PROGRAMME

The multidisciplinary bug catches some people early. Gericke Sommerville, who is now a postdoc in an IGERT programme for interdisciplinary mathematics, ecology and statistics at Colorado State University (CSU) in Fort Collins, studied mathematics and biology as an undergraduate, quantitative ecology and ecological monitoring as a graduate student and now teaches graduate students quantitative methods for biology. And chemist Jennifer Tripp, who has an NSF international postdoc award, combines her training and interest in chemistry and archaeology in a fellowship in the radiocarbon-dating unit in the archaeology department at Oxford University.

Some scientists start on multidisciplinary programmes in their graduate studies. Norbert Lanfer, a postdoc in the Urban Ecology Research Training Group at Humboldt University, Berlin, is a geographer by training. But his PhD research in Ecuador was on the sustainability and development of small farmers in the Amazon. Now he is involved in a research project on the growth conditions for the praying mantis in northern Europe. The insect is native to the Mediterranean, and has become an invader species in Berlin.

But for many, going multidisciplinary after your PhD is the best time to make the switch. Jeffrey Achter, who has just begun a postdoc at CSU, has a background in pure mathematics. But he wanted to make a shift to "something more connected with the rest of the world", so he is applying his mathematical skills to modelling animal population dynamics.

Zhijie Tan, a postdoc in the multidisciplinary physics programme at the University of Missouri, Columbia, chose that route because he sees pure physics as a career dead end: "There will be few significant findings in the field of pure physics in the coming decade or so," Tan says. He sees more opportunities in structural biology, and is becoming a biophysicist, studying the structure of nucleic acids in solution.

FINDING A JOB

Many postdocs who follow the multidisciplinary route see career advantages. Bradford Percy, who has a wide background in mathematical biology, is working on modelling gene networks in a new VIGRE programme in the department of computational and applied

interdisciplinary research and collaborations, but he also immediately became involved in an application to NSF for an IGERT grant for the department to carry out multidisciplinary training.

Academia is not the only option after completing a multidisciplinary postdoc. Graduates and postdocs from the Wind Science and Engineering Research Center at Texas Tech University, Lubbock, "have a variety of places they can go", says Kishor Mehta, the principal investigator on Texas Tech's IGERT programme in wind science and engineering. Some do enter academia, but others have gone on to work for consultancies, the insurance industry, the National Oceanic and Atmospheric Administration (NOAA) and the Federal Emergency Management Agency.

Some have an entrepreneurial bent. Postdoc Gleb Yushin, who works on nanoscale engineering in the Drexel University IGERT programme, sees materials science as a good springboard. "If I feel that some newly developed technology is asking to be commercialized and there is nobody to push it, I will start a company," he says.

Many US postdocs come from overseas and plan to return home after completing their fellowships. Balazs Hegedus, who is studying the viscoelastic and physical cell-adhesion properties of three-dimensional aggregates of tumour cells in the biological physics research group at the University of Missouri, Columbia, plans to return to his native Hungary to work in the tissue culture laboratory of the National Institute of Neurosurgery in Budapest. "This training here might be a basis for a long-term collaboration and a good opportunity to study new approaches while doing research," he says.

Despite the opportunities, some of those taking the multidisciplinary path worry about their future career. David Tallmon, an NSF-funded postdoc at Joseph Fourier University in Grenoble, France, is concerned that his training, which spans computation, statistics, biology and ecology, will not fit a niche — especially in academia. "If I go the academic route, there is little to no reward for being multidisciplinary in one's focus and working directly on conservation problems," he says.

But Metters thinks the risks are more than rewarded when one starts seeing connections "that no one else can see". He'd do it again like a shot, he says. ■

Myrna Watanabe is a freelance science writer based in Patterson, New York.

Web links

- SPRING (Strategic Partnership for Research in Nanotechnology)
♦ www.cm.utexas.edu/SPRING/index.htm
- The Urban Ecology Project
♦ www.urbanecology.washington.edu
Institute of Agricultural and Urban Ecology Projects
- ♦ www.agrar.hu-berlin.de/ASP/ASP.htm
Boston University Bioinformatics Program
- ♦ bioinfo.bu.edu/index.shtml
Max Planck Institute for Mathematics in the Sciences
- ♦ www.mis.mpg.de
Bioinformatics Center, Institute for Chemical Research, Kyoto University
- ♦ <http://www.bic.kyoto-u.ac.jp/egis>